Molecular engineering of the catalytic antibody 20G9

Smith, Lee C.

Award date:
1999

Awarding institution:
University of Bath

MOLECULAR ENGINEERING OF THE CATALYTIC ANTIBODY 20G9

submitted by Lee C. Smith

for the degree of PhD

of the University of Bath

1999

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

20G9 is a murine IgG that catalyses the hydrolysis of phenyl acetate to phenol and acetic acid. The antibody was raised to a phenyl phosphonate hapten, which was designed as a stable analogue of the assumed tetrahedral transition state formed during the hydrolysis of phenyl acetate. Site directed mutagenesis has been used to investigate the mechanism of the catalytic antibody 20G9. The ultimate aim of the research is to try to understand the important features of catalytic antibodies so that rational alterations can be made to improve antibody mediated catalysis.

The antibody combining site of 20G9 was modelled using the antibody modelling program, *AbM*. An attempt was made to identify potential key catalytic residues based upon proximity, previous kinetic characterisation of the antibody and residue function found in other catalytic antibodies. Residues were selected and PCR site-directed mutagenesis performed which succeeded in producing five mutant antibodies each containing single residue changes in the combining site as determined by sequencing. Expression was attempted in three different plasmid vectors of which pUC119His6*mycXba* was chosen. The mutant inserts were cloned into this vector and expression optimised. Recombinant scFv antibody was expressed in *Escherichia coli* strain TOPP2, grown at 30°C and induced for 24 hours with 1mM IPTG at 26°C. The scFv antibodies were identified by western blotting against the *c-myc* tag using the antibody 9E10 and purified by IMAC using a c-terminus histidine tag. Only the mutant $Y^{L}96A$ failed to produce sufficient material for kinetic analysis. Kinetic characterisation of the 20G9 scFv saw similar activity displayed to that of the hybridoma derived 20G9. When kinetic characterisation was performed upon the four 20G9 scFv mutants, $R^{H}50A$, $R^{H}56A$ and $Y^{L}94A$ were seen to be active. $R^{H}50A$ displayed activity comparable to the wild type 20G9 scFv. With $R^{H}56A$ and $Y^{L}94A$, the K_M during the burst phase increased two-fold which may indicate involvement in binding of the substrate. No measurable activity was seen for the mutant $Y^{L}91A$ indicating that a residue critical for activity had been eliminated. It is likely that this residue forms an acetyltyrosyl intermediate during the catalytic hydrolysis of phenyl acetate.

Acknowledgements

Many thanks to Jeff Gyi for his practical supervision during the first eighteen months of this project and to Steve Searle with his computer assistance in the modelling of 20G9 and help with UNIX, Molscript and Insight II. Thanks to Sue Phillips and Alison Jones for their assistance with the 'frustrating' molecular biology that was often encountered during this project. Also, much thanks to Rob Griest with his valuable lessons in the use of FPLC systems and Western Blotting. Thanks to Jonathan Cox for his help in the analysis of transition state analogue. A special thanks to Rachel McCarthy and Simon Palmer for the humour that kept me smiling during the more difficult moments of the project.

Much thanks to my supervisor Tony Rees for presenting the opportunity to work on this project and for his help in producing this thesis.

Finally I must thank my wife, Annamarie for her enormous patience and un-relenting support over the last five years.

Lee C. Smith, April 1999.

ABBREVIATIONS

GENERAL

Å	**Å**ngstrom unit, 1**Å** = 0.1nM
Ab	antibody
AbM	Antibody modeller
ACS	antibody combining site
CAMAL	Combined antibody modelling algorithm
ccc	covalently closed circular
cDNA	cloned deoxyribonucleic acid
CDR	complementarity determining region
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	2' deoxy nucleoside triphosphate
ds	double stranded
FPLC	flow pressure liquid chromatography
HPLC	high pressure liquid chromatography
Ig	Immunoglobulin
K_a	Association constant
K_d	Dissociation constant
k_{cat}	Turnover constant
K_M	Michaelis constant
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
LMP	Low melting point
mAb	monoclonal antibody
MCS	Multiple cloning site
MHC	Major histocompatibility complex
mRNA	messenger ribonucleic acid
NMR	nuclear magnetic resonance
PDB	Protein data bank
RBS	Ribosome binding site
RNA	ribonucleic acid
scFv	single chain variable fragment
SCR	Structurally conserved regions
ss	single stranded
TCR	T-cell receptor
TSA	Transition state analogue
Tm	melting temperature
tRNA	transfer ribonucleic acid
UV	ultraviolet
w/v	% weight by volume
v/v	% volume by volume
V_{max}	Maximum reaction velocity
V	Reaction velocity

WT	wild type

CHEMICALS

AMPS	Ammonium persulphate
BSA	bovine serum albumin
DMSO	dimethyl sulphoxide
EDTA	diamino ethanetetra-acetic acid
MOPS	3-(N-morpholino) propane sulphonic acid
PBS	phosphate buffered saline
SDS	sodium dodecyl sulphate
TBE	Tris borate EDTA
TEMED	N, N, N', N'-tetramethylethylene diamine
Tris	Tris (hydroxymethyl) amino methane
Tween-20	Polyoxyethylene sorbitan monolaurate

AMINO ACIDS

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

CHAPTER 1

1. INTRODUCTION

This thesis concerns the modelling, mutagenesis, expression and characterisation of the antibody 20G9 and its mutants. This introduction will briefly cover the organisation of antibodies at a genetic and structural level. In addition it will give an overview of catalysis by antibodies and the methods by which they can be rationally and randomly engineered for expression and characterisation.

1.1 THE IMMUNE RESPONSE

Immunity is a word derived from the Latin *immunitas* which was given to Roman senators during their office tenure allowing exemption from various civic duties and legal prosecution. Today, immunity stills means protection but within a clinical context indicates protection from infectious disease. Immunity has gradually evolved within animals enabling them to recognise and destroy a wide variety of pathogens that may harm them. The immune system of mammals is currently the most widely understood and is increasingly manipulated by both the pharmaceutical and biotechnology industries for many clinical and scientific applications.

The ability to specifically recognise molecules on pathogens is key to the immune system. There are two tiers to the immune system:

i) natural immunity is the body's first line of defence and is non-discriminatory in nature. This includes physiochemical barriers such as skin, mucus membranes and circulating molecules such as complement and phagocytic cells. The levels tend to remain static within the body and are not enhanced by repeated exposure to foreign bodies.

ii) acquired immunity is a defence mechanism that is induced and enhanced by repeated exposure to foreign particles. The predominant cell types involved in this

immune response are the T and B lymphocytes both of which are responsible for specific recognition of potential pathogens.

The B-lymphocytes possess membrane bound receptor molecules upon their surface but are also able to express soluble antibody molecules identical in terms of specificity to those bound to the membrane. Both molecule types are able to bind to native conformations of soluble or membrane bound antigen but only antigen bound to the membrane bound receptor molecules results in clonal expansion. This causes naive B cells, specific for that antigen, to differentiate into plasma B cells capable of secreting large amounts of antibody (**Ab**) of equal specificity. There also follows reduced differentiation into memory B cells. The high amounts of secreted antibody trigger a cascade effect, resulting in activation of the natural immune functions such as complement and stimulation of macrophage, natural killer cells, etc.

However, the B cells are not alone in the immunisation process and their expansion is closely linked to the T cell lymphocytes. The T cells also possess membrane bound receptors from the immunoglobulin superfamily whose functional role and genetic origins are distinct from their antibody cousins. These are T-cell receptors (TCR) and they recognise antigens that have been endocytosed by antigen presenting cells (APC's) such as B cells and macrophage. Within the B-cell the internalised proteins are degraded and subsequently bind and complex with major histocompatibility complex class II molecules (MHC II) found within the cells rapidly followed by their presentation upon the surface. TCR molecules upon the T-cells recognise the processed antigen bound to the MHC molecule along with other simultaneous interactions between accessory molecules, i.e. CD4. This triggers T cell activation, which in turn induces or suppresses other aspects of the immune cascade by cytokine release, including differentiation and proliferation of the antibody producing B cells (Figure 1.1).

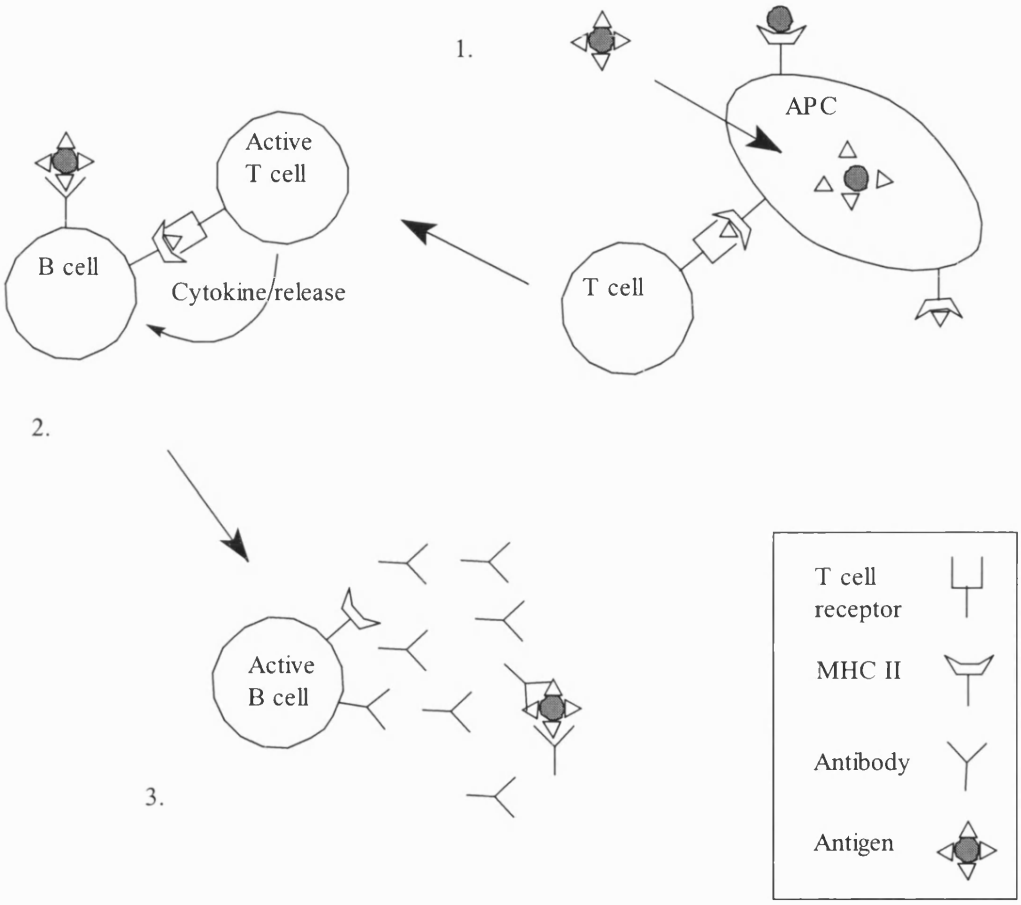


Figure 1.1 A widely accepted version of the interaction between the B and T cell lymphocytes and accessory cells. Antigen is internalised by the accessory cells, also named antigen presenting cells (APC), i.e. macrophage, which fragment the foreign body, complex its various components with MHC class II molecules and display them on the cell surface. The T cell binds the APC and its associated displayed complexes using its cell receptors and adhesion molecules and is subsequently triggered into cytokine production. This in turn stimulates the B cells into clonal expansion and high level expression of antibody molecules specific for a single epitope.

1.2 ANTIBODY STRUCTURE

1.2.1 The Immunoglobulin Structure

An antibody molecule has two functions; to tightly bind to its epitope and to stimulate complement release and immune cell activation. These functions are combined in a single multidomain immunglobulin (Ig) molecule of which five Ig classes exist; IgA, IgD, IgE, IgG and IgM, each possessing different structures and functions. The IgG molecule is a tetrameric molecule consisting of two identical 25kDa light chain polypeptides (V_L and C_L) and two identical 50kDa heavy chain polypeptides (V_H, C_H1, C_H2 and C_H3) linked by disulphide bonds (Figure 1.2). It represents 70-75% of immunoglobulin found in mammalian serum and so is by far the most common antibody class found in the circulation.

IgA normally accounts for 15-20% of total immunoglobulin found in human serum but is the principle immunoglobulin found in external secretions, i.e. saliva, tears, breast milk, respiratory secretion. IgA essentially has the same structure as IgG, two light chain molecules (V_L and C_L) and two heavy chain molecules (V_H, C_H1, C_H2 and C_H3) but can exist as a monomer, a dimer that is held together by a J-piece (a 15kD polypeptide that is disulphide bonded to the F_c) and also as multimers, which are joined by disulphide bonding.

The overall structure of IgD is the same as IgG except that is possesses an extra constant domain on the heavy chain (V_H, C_H1, C_H2, C_H3 and C_H4). These immunoglobulins are normally present in human antisera at very low levels, < 0.2% with the role of this immunoglobulin poorly understood. IgE is the antibody class that mediates certain allergic reactions and is involved in anaphylatic shock. They are similar in size to IgD, possessing an extra constant domain (V_H, C_H1, C_H2, C_H3 and C_H4) and are normally present in trace amounts within sera.

IgM is the final class of immunoglobulin and was the first to be discovered. The monomer of this immunoglobulin is similar in size to IgD and IgE in that it has an extra constant domain (V_H, C_H1, C_H2, C_H3 and C_H4) but it exists as a pentamer in

the serum. The pentameric immunoglobulin is linked by disulphide bonds on the C_{H3} and C_{H4} domains and also by a J-piece similar to that used to join IgA dimers. This antibody class is the first antibody to be produced during an immune response, but is soon eclipsed in concentration by IgG as the response progresses. Subsequently, IgG represents approximately 70% of total immunoglobulin seen in human serum. This thesis concerns the engineering of a monoclonal IgG and further discussion will concentrate on this class of antibody.

The Ig molecule can be divided into several smaller function based units; the F_{ab} (antigen binding fragment), which has just a single binding site and the F_c (fragment crystallisable) which possesses cell effector functions. In addition the F_{ab} can be further divided into a F_v (variable fragment) which has no constant domains but still retains the antigen binding function (Figure 1.2).

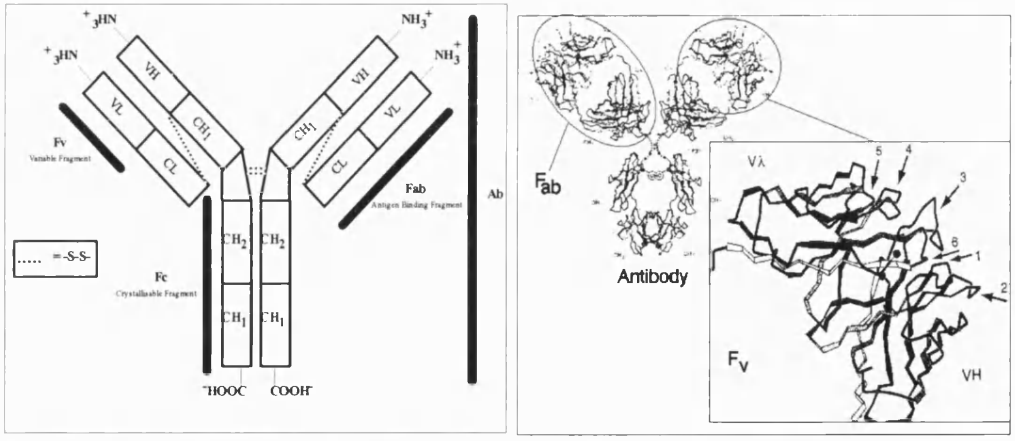


Figure 1.2 Schematic structure of an IgG molecule. The molecule is tetrameric consisting of two 25kDa light chains and two 50kDa chains joined together by disulphide bonds. The molecule can be divided into 3 main sections; F_{ab} (antibody binding fragment), F_c (crystallisable fragment) and F_V (variable fragment). The F_V contains 6 CDR regions that form the antigen binding region of the molecule.

The light chain is composed of two domains, V_L (light variable) and C_L (light constant), whereas the larger H chain is composed of 4 domains, V_H, C_H1, C_H2, C_H3 (or 5 if the C_H4 domain found in IgD, IgE and IgM classes is included). The variable domains (**V**) contain the hypervariable regions that are held in place by conserved framework regions. The constant domains (**C**) are highly conserved and form the antiparallel β-barrel structures that are seen across all antibody classes. A common feature of all Ig are three conserved cysteine residues, found in the hinge region, which are necessary for the formation of disulphide bridges between the various domains and are vital for the formation of the quaternary structure [Lesk and Chothia, 1982].

1.2.2 The antibody combining site

Several three-dimensional structures of whole antibodies have been solved by X-ray crystallography [Chang *et al*, 1994; Lescar *et al*, 1993; Davies & Padlan, 1990, Harris *et al*, 1995]. In addition, structures have been solved of F_{ab} [Braden *et al*, 1994; Griest *et al*, 1992; Poljak, 1973] (in its natural and antigen complexed states) as well as the F_c [Diesenhofer, 1976]. This has allowed full structural characterisation of antibody molecules giving much insight into their folding. The V_L and V_H domains each contain three hypervariable loops termed complementary determining regions (CDR's). The six CDRs pack together to form the antibody-combining site. The sequences between each CDR are termed framework regions. Each framework segment is a β-strand that folds with other strands to form a β-sandwich which, when combined with the β-sandwich of the opposing chain, forms a β–barrel structure.

Examining the constant and variable domains, the constant domains consist of seven β-strands that form two β-sheets which, joined by disulphide bridges form an antiparallel β-barrel. The variable domains are similar but consist of nine β-strands packing similarly to the constant region forming an antiparallel β-sheet. It is these loops joining the variable strands that are hypervariable and are denoted CDR's 1, 2 & 3, while the constant domain remains largely conserved. Although the CDR

loops are hypervariable, the β–barrel packing of the V domain is normally unaffected. This stability can be exploited and CDR loops have been transplanted from one antibody to another with the integrity of the β-parallel barrel usually remaining intact. This has allowed CDR's from a mouse antibody to be grafted onto framework regions of human antibodies allowing clinical use in humans [Riechman *et al.,* 1988]. Occasionally, sequences differences between human and murine frameworks have necessitated the grafting of additional, non-CDR residues, to maintain the correct structure [Foote & Winter, 1992].

The specificity of the antibody combining site is mostly determined by the CDRs but can also be affected by the framework regions, though usually restricted to the outer fringes of the F_v regions [Ward *et al*, 1997; Winter & Foote, 1992]. A few residues at specific sites within the loops determine the backbone conformations of the six CDRs. These residues have been found to be conserved in different V_L and V_H sequences and are described as 'key' residues, the existence of which has allowed the definition of a group of homologous 'canonical' structures. For the CDRs to maintain their specific conformation, these residues must be conserved. The key residues that are associated with the various canonical structures can be used to predict any canonical class on any given antibody with the exception of the H3 loop [Morea *et al*, 1998; Al-Lazikani *et al*, 1997; Barre *et al*, 1994; Chothia, 1989; Chothia & Lesk, 1987].

1.2.3 Antibody-Antigen Interactions

Antibodies are molecules that bind tightly to immunogenic molecules, or antigens to which the B-cells have been exposed. Before a tight association between two molecules can form, large entropic barriers must be overcome. Where there is loss of free rotation of mobile elements such as side chains within the macromolecules, there is also a loss of entropy. Contributions to overcome these barriers derive from van der Waals interactions, hydrogen bonds and the salt bridges formed between the antibody epitopes and its antigenic determinants upon the surface of the antigen. These alone are insufficient to cause tight binding, and shape complementarity between the surfaces must be sufficient to drive the association. Where this

complementarity is not perfect, water molecules can assist in binding; several studies have shown water involvement at the interface of the two molecules [Tulip *et al.*, 1992; Bhat *et al.*, 1994] contributing up to 25% of the total binding strength [Corell & Mallqvist, 1997].

1.3 GENERATION OF ANTIBODY DIVERSITY

In animals, antibodies are one of the major tools used to defend the body against pathogens. In order to recognise the almost infinite range of pathogens to which the immune system will be exposed, the immune system has had to evolve to respond and produce an almost equally diverse repertoire of antibodies used to recognise and remove foreign antigens from the blood stream.

Production of antibodies is a constantly evolving process. B cells producing antibodies to combat an initial exposure to an antigen will not be expressing an identical antibody a few days later. The antibody will have evolved and developed an improved affinity towards the antigen and consequently will be more effective at lower quantities. The lower affinity antibodies produced initially are the result of a primary response of the naive population of lymphocytes and result in plasma and memory cells. The secondary response produces higher affinity antibodies and is due to the response using the available pool of memory cells that differentiated during the primary response. In between these two responses the affinity of the antibodies towards the antigen improves and is due to a combination of somatic mutation of existing antibodies and selection of antibodies with gene rearrangements that confer improved binding characteristics towards the antigen.

1.3.1 Immunoglobulin gene organisation and rearrangement

Unlinked gene families encode the κ and λ light chains and the α, β, ε, γ and μ heavy chains, with the genes for each of the chains found on separate chromosomes. The variable and constant segments for each chain are however proximal to one another [Tonegawa, 1983; Calabi & Neuberger, 1987]. The immunglobulin heavy gene locus is composed of 4 gene segments. The 5' end of the gene is comprised of

multiple V exons followed by D (diversity) exons and J (Joining) exons. The 3' end of the gene is comprised of the C (constant) segments with each segment (or exon) separated by non-coding DNA that is eventually discarded through gene rearrangement and transcription. Rearrangement of heavy chains brings the exons together forming a VDJC gene while in light chains a VJC gene is formed (Figure 1.3). When the various exons are spliced together, an enhancer and a promotor, which are distal from one another before recombination, are brought into close proximity triggering transcription of the gene.

If the variable chains of each of the heavy and light chains are examined, the V segments encode the final four residues of a peptide leader sequence and the first 95 amino acids of the N-terminus residues that form the CDR 1, 2 and framework regions 1-3. In the heavy chain, the D segments and the J segments contribute to CDR3 and the fourth and final framework region whereas in the light chain the CDR3 and the fourth framework region are comprised of only the J segment (Figure 1.4). Association of these elements is thought to generate around 5.0×10^4 sequences in the heavy chain, while there are only approximately 1.0×10^3 sequences in the light chains due to the lack of the diversity segments (Figure 1.4). Consequently the CDR3 of the heavy chain has significantly more variability than that of the light chain CDR3. In addition, the L2 CDR is often found not to bind to small antigens even when all loops are found to be in contact [Tormo *et al.*, 1994], suggesting a somewhat smaller contribution of the light chain to combining site variability.

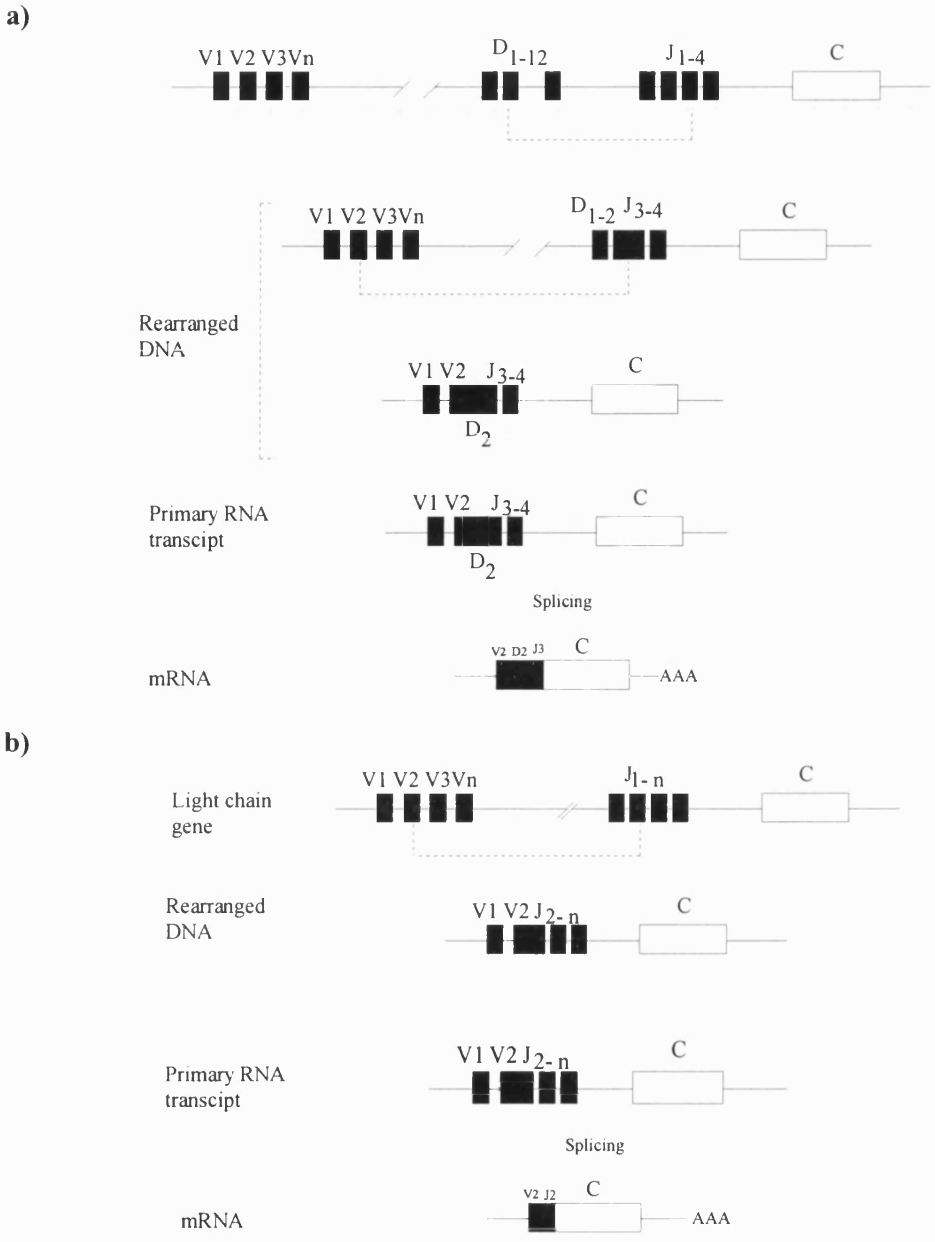


Figure 1.3 a) immunoglobulin heavy chain germlines and their subsequent rearrangement into the variable regions of the V and C domains of the antibody. b) immunoglobulin light chain germlines and their subsequent rearrangement into the variable regions of the V and C domains of the antibody.

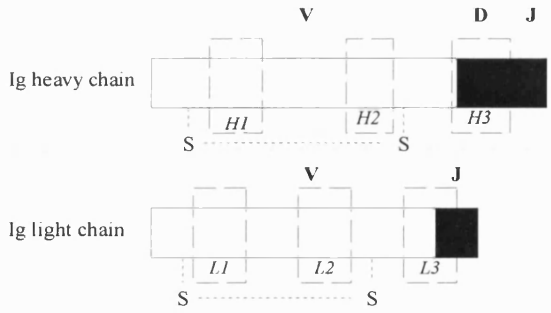


Figure 1.4 How the immunoglobulin genes relate to the domains of the polypeptide chains. The dashed boxes indicate the CDRs displaying the genes that they are formed from. The L_3 CDR is composed of both V-J genes where as the H_3 CDR is composed of V-D-J genes and consequently is the most variable of the hypervariable regions.

Site specific recombination is achieved using an array of enzymes collectively termed recombinases that are involved in the formation of complete variable regions. The enzymes are thought to recognise specific nucleotide recognition sequences found in the 3' DNA of each V segment and at the 5' of each J segment as well as sequences found in both sides of each D segment (where present). The DNA recognition sites are conserved nucleotide heptamers or nonamers separated by non-conserved 12 or 23 base pair (bp) nucleotide spacers. The adjacent heptamers/nonamers on the V segments recognise complementary sequences on the J segments. However it is the 12/23bp spacers which dictate joining as a 23bp spacer must face a 12bp spacer on the opposing sequence and *vice versa*. This mechanism is thought to be instrumental in preventing V segments from combining together. Heptamer/nonamer recognition allows the formation of DNA loop structures with the intervening DNA excised, followed by joining of the cut ends. It is at this point where non-germline diversity is introduced into the antibody sequences due to imprecise restriction and ligation of the ends, described as *junctional diversity* (heavy chain only). As a result, an identical set of gene segments can result in different amino acid sequences following recombination. In some cases this may lead to non-functional recombinations that are out of frame. This is sometimes overcome by the B cells deleting nucleotides from the non-functional gene upstream of the join resulting in a functional sequence. Additional genetic diversity is added to the immunoglobulin rearrangement by V_H exchange [Kleinfield *et al*, 1986] and through N-region diversity (heavy chains only) where

insertion of nucleotides at the V-D and D-J junctions during rearrangement of the heavy chain genes occurs. These nucleotide additions are referred to as N segments and may extend with as many as 30 nucleotides that sometimes result in a much extended CDR3 loop.

After immunoglobulin gene rearrangement has taken place, the B-cells expressing low affinity antibodies undergo a mutation and selection process termed somatic hypermutation [French *et al.*, 1989; Kocks & Rajowsky, 1989]. Mutations occur at various positions throughout the variable regions of heavy and light chains, which may have an effect upon the affinity of the antibody towards its antigen. The diversity introduced by somatic hypermutation is estimated to increase the germ-line repertoire from 10^9 to 10^{30} different binding determinants [Winter & Milstein, 1991]. Mutations are most common in the hypervariable regions and normally result in an increase in the affinity of an antibody, although the inverse is also true [Tonegawa, 1983; Wysock *et al.*, 1990]. The higher affinity antibodies are subsequently selected, followed by clonal expansion of the B-cell germline. In reality, this huge range of generated antibodies are not produced indefinitely, otherwise huge amounts of antibody would be present at any one time in the circulation. To overcome this, memory B-cells exist which allow rapid upregulation, when necessary, of antibody numbers specific for a particular antigen.

1.4 EX-VIVO EXPRESSION OF ANTIBODIES AND THEIR DERIVATIVES

1.4.1 Production of whole antibodies

Many of the antibodies that are found in blood differ from one another with regard to their class, isotype and in their specificity towards a given antigen. However the antibodies are all part of an immunoglobulin superfamily and are sufficiently similar in that they can all be easily isolated from blood products by similar methods. Blood, which has been separated from its main cellular components, contains this mix of antibodies and is called serum. If the antibodies within the serum are purified using, for example protein A, the purified antibodies are called *polyclonal antibodies* as they contain a heterologous population of antibodies which

have different target epitopes and different binding specificities derived from different B cell clones.

It was not until the mid 1980's that a homogenous population of 'monoclonal' antibodies could be produced which targeted the same antigen with identical specificity, possibly the most significant immunological and biotechnological event since the introduction of vaccines. The breakthrough in the production of antibodies was achieved through the use of myeloma cells. These are immortal tumour cells derived from antibody-secreting plasma cells and secrete high numbers of a single antibody and allow isolation of a single cell providing a limitless source of antibody, each a clone of the other. Particular myeloma clones were selected that retain the immortal phenotype but had lost their antibody expression. They also contained certain genetic features that enabled selection in a mixed population of myeloma cells, fused myeloma-B cell and B cells. The introduction of the antibody genes is achieved by cell fusion between the myeloma cells and the B cells from the spleen. The spleen cells that were exposed to antigen to raise specific antibodies are fused with the myeloma cells using polyethylene glycol that forces the cell membranes to blend subsequently leading to hybrid cells named *hybridomas.* These cells possess the characteristic immortality of the tumour cells but retain the antibody genes of the spleen cells. The cell line producing the antibody of desired specificity can be separated from the other cells using limit dilution and detected using enzyme linked immunosorbent assays (ELISA). Once selected, they can be cloned to establish permanent cell lines from which the monoclonal antibodies may be obtained. Repeated immunisation of mice with an antigen enables the isolation of high affinity, highly specific, monoclonal antibodies. The production of monoclonal antibodies by the routine procedure outlined above is a process which takes approximately six months, although more rapid methods of monoclonal antibody production have been developed such as direct immunisation of spleen cell culture [Stahl *et al.*, 1995] and phage display (described in 1.4.3). At present, the major flaw of these non-classical methods is the absence of an effective somatic hypermutation stage, although artificial diversity has been introduced using phage

antibodies by methods such as chain shuffling [Collet *et al*, 1992] and CDR randomisation [Barbas *et al*, 1992]

1.4.2 Expressing recombinant antibody fragments

The F_c effector functions of whole antibody are surplus to requirements if only specific binding of the F_{ab} is required. Often the large size of the whole antibody can restrict their use in therapy due to poor tumour penetration. Although whole antibodies have the advantage that they are slowly depleted from circulation, they provide a much larger antigenic surface to which host antibodies may be raised, leading to undesirable immunogenic side effects. However the use of antibody fragments reduces or eliminates these problems and they have a considerably shorter half life within the body [Colcher *et al*, 1990]. Antibodies or antibody fragments have been successfully produced in bacteria, [Field *et al*, 1989; Mamalaki *et al*, 1993]; *E. coli* [Pluckthun, 1991], *Bacillus subtilis* [Wu *et al*, 1993], yeast [Bowdish *et al*, 1991], insect cells [Hasemann and Capra, 1991], mammalian cells [Bebbington, 1991; Gilliand *et al.*, 1996] and plants [Owen *et al*, 1992; Tavladoraki *et al*, 1993; Benvennto *et al*, 1995]. Many laboratories express scFv, dsFv [Reiter & Pastan, 1998] and Fab fragments of antibodies in *E. coli* (Figure 1.5). However the associated problems of expressing eukaryotic proteins within prokaryotes has led to the development of eukaryotic expression hosts such as yeast and mammalian cells, although these systems present their own problems.

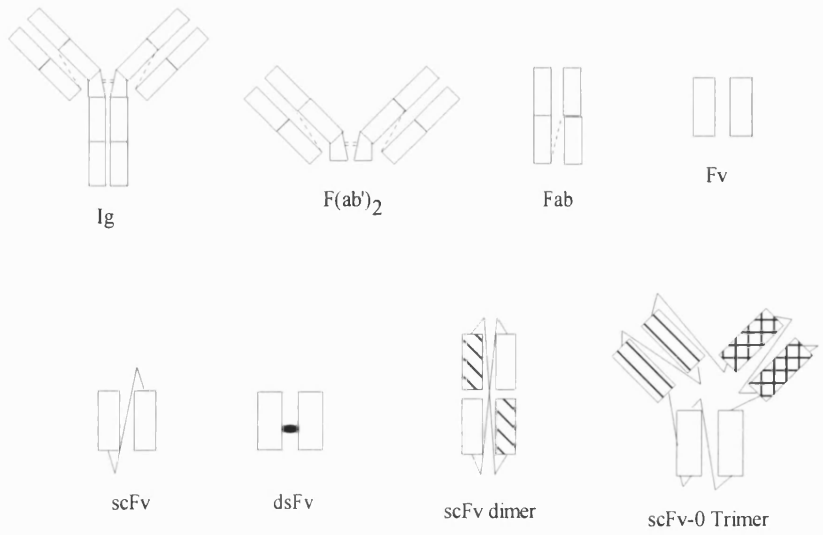


Figure 1.5 Schematic diagrams of antibody fragments: IgG, Fab, $F(ab')_2$, Fv, scFv, dsFv, scFv dimer and scFv-0 residue trimer. Some of the above fragments ($F(ab')_2$ and F_{ab}) can be obtained by enzymatic digestion of whole immunoglobulins using pepsin and papain while others can be expressed in bacteria, yeast, mammalian cells and plant tissue although the antibody fragments are sometimes toxic towards the host cell.

Expression of antibodies in prokaryotic systems has a number of problems. For example, not all codons commonly found in eukaryotic genes translate well in *E. coli.* This is possibly due to variation in the levels of specific tRNAs used in translation where few or no tRNAs exist for specific codons [Ikemura, 1981]. Consequently these codons are rarely present within *E. coli* genes and attempts to express eukaryotic genes containing these codons can present problems. Again, some antibody sequences appear toxic when expressed in bacteria, possibly due to overloading of the cell's transport machinery. This can result in difficulty in the expression of particular antibodies and can sometimes result in bacterial cell lysis or in degradation of the expressed protein [Pluckthun, 1994; Darveau *et al*, 1992].

The vectors used in *E. coli* can make an important difference to the successful expression of recombinant antibodies: vector stability and promotor efficiency may both be critical for success [Bell, 1995]. Some vectors are unstable within *E. coli,*

so that even when antibiotic markers are present the vectors can suffer large deletions of nucleotide sequences. Some commercially available vectors such as pKK233-2 have exhibited this problem. Therefore, it maybe important to transform and survey many vectors within *E. coli* in order to select a suitable and robust expression system for a particular antibody fragment. As mentioned previously, recombinant antibodies can often be toxic towards their bacterial hosts [Pluckthun, 1994] and this, in combination with pressure on the bacterial cell to produce high levels of protein, can cause plasmid loss and subsequent loss of expression. It is critical to have a built in method of controlling expression at all times, such as via an inducer which can reduce the likelihood of plasmid damage or loss. However, the use of strong promotors to generate high expression levels can sometimes result in 'leaking' where low level expression occurs in the absence of the inducer. Additional controls have been developed to prevent this using inhibitors such as *Lac*I that are present in some vectors [Miller & Albertini, 1983; Fraipont *et al*, 1994]. When using *Lac*I inhibitors, the addition of glucose can be used to suppress expression and this, allied with inducers, can lead to a well regulated, high level expression system [Tagami *et al*, 1995; Ishizuka *et al*, 1994]. Examples of promotors that promote strongly, yet are tightly regulated, are P_{lac}, P_{tac}, P_{tnc} [Trukhan *et al*, 1988; Amann *et al*, 1988].

1.4.3 Phage display

The use of λ-bacteriophage has allowed antibodies to be produced *in vitro* avoiding the time consuming classical method of monoclonal antibody production while utilising a combinatorial method which improves the specificity and affinity of the antibody molecule. It is the use of the polymerase chain reaction (PCR) which has enabled rapid reproduction of antibody gene repertoires through the use of degenerate oligonucleotide primers. These primers include a random nucleotide on 'the wobble' while retaining the relatively conserved positions seen within the antibody sequences. The primers can also be designed to include restriction sites which allow the immunoglobuin genes to be amplified and easily transferred into DNA vectors [Orlandi *et al.,* 1989; Gussow *et al.*, 1989]. Genetic manipulation to fuse scFv or F_{ab} genetic sequences with the N-terminus of phage minor coat protein pIII or gene VIII [McCafferty *et al.*, 1990] has allowed antibody fragments to be

displayed on the bacteriophage surface. These correctly folded fragments are incorporated into the phage coat allowing binding of antigen and are termed *phage-antibodies*. The recombinant phage-antibody contains the DNA of the displayed antibody so that it can be 'captured' based on its binding properties towards an antigen bound to a solid surface and consequently the gene for the antibody isolated. The oligonucleotides generated contain the random nucleotide at 'the wobble' within the oligonucleotide primers so that the V genes of the antibody are amplified. The antibody gene repertoires are then amplified using PCR primers and subsequently cloned into the phage, creating a phage library that displays between 10^6 and 10^9 antibodies of different specificities. Antibodies displaying the desired binding properties can be selected on a particular antigen while those failing to bind eliminated through washing [Winter *et al.*, 1994; Griffiths *et al.*, 1994]. A major benefit of this system is that human antibodies can be easily generated [Persson *et al.*, 1991] avoiding the standard re-shaping protocols in order to avoid causing adverse immune responses [Gorman & Clark, 1990].

Phage display technology has progressed most recently with the selective infective phage technology (SIP). The phagemid displaying its antibody is non-infectious until an interaction occurs between the antibody and an antigen within the bacterial cell periplasmic space. This triggers cell apoptosis releasing large amounts of infectious phage particles [Spada *et al.*, 1997].

1.5 BIOLOGICAL CATALYSTS

1.5.1 Bioenergetics and catalysis

Enzymes are biological catalysts that are thought to have slowly evolved over millions of years with many achieving a state of kinetic perfection [Fersht *et al.*, 1985]. Kinetic perfection exists when the catalytic turnover (k_{cat}) of substrate to product by the enzyme is faster than the rate of substrate diffusion into the enzyme's active site (K_M). This means that the rate of catalysis is *diffusion limited* where diffusion of the substrate into the active site limits the rate or turnover of the enzyme ($V = k_{cat}/K_M$), so that it cannot be higher than 10^8-10^9 M^{-1} s^{-1}. Most enzymes accelerate reactions are at least in the order of 10^6 M^{-1} s^{-1} although enzymes such as carbonic anhydrase, acetylcholinesterase and triosephosphate

isomerase are between 10^8-10^9 M^{-1} s^{-1} and so are among those considered to be the most kinetically perfect [Albery & Knowles, 1977; Fersht, 1985]. It is important to understand that an enzyme is a catalyst and cannot alter the equilibrium of a chemical reaction but will only accelerate the attainment of equilibrium.

When a chemical reaction occurs, in order for the substrate (**S**) to form the product (**P**), it must pass through a transition state ($\mathbf{S}^{\Psi}$) that has a higher Gibbs free energy than either S or P. The difference in free energy between the S and S^{Ψ} is the Gibbs free energy of activation (ΔG^{Ψ}). If this value is reduced the attainment of equilibrium is accelerated (Figure 1.6). Enzymes stabilise the reaction transition state thus reducing ΔG^{Ψ}, and consequently accelerating the attainment of equilibrium. With most enzyme catalysed reactions, the catalytic rate (V) varies with substrate concentration (S). At a fixed enzyme concentration, V is directly proportional to S when at small concentrations, whereas V is almost independent of S when at high levels and the enzyme is saturated. Based upon the Michaelis-Menton model [Palmer, 1995], V_{max} is the resulting rate when the enzyme is fully saturated with substrate. The K_M (the Michaelis constant) is the substrate concentration [S] at which the reaction rate is half of the maximal rate ($V_{max}/2$) (Figure 1.7).

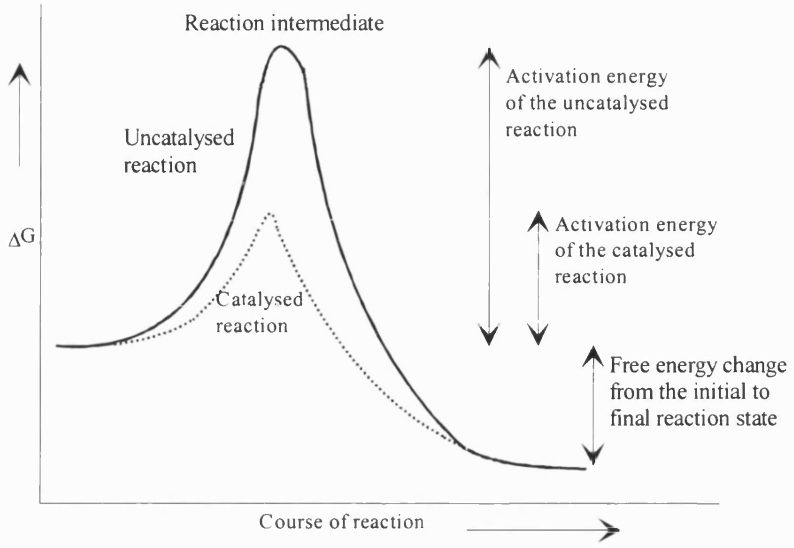


Figure 1.6 Reactions are accelerated by enzymes decreasing the $\Delta G^{\ddagger}$.and thus the reaction is accelerated. The kinetic profiles demonstrate catalysed and uncatalysed reactions.

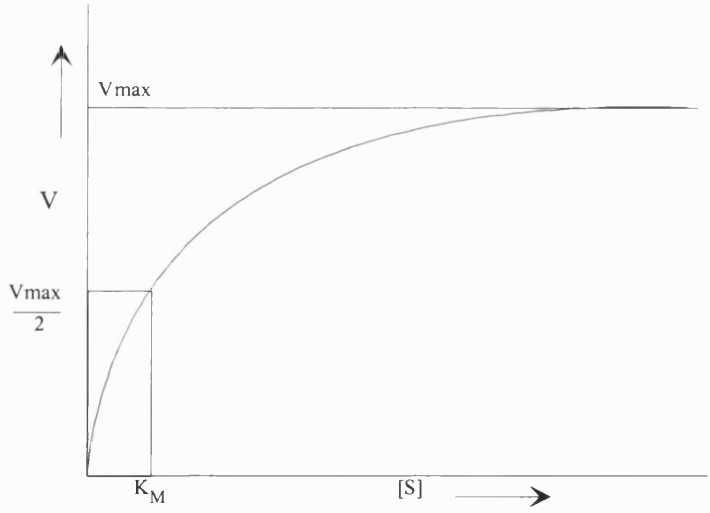


Figure 1.7 The reaction velocity V is linked to substrate concentration [S] for an enzyme that obeys Michaelis-Menton kinetics. V_{max} is the maximal reaction velocity and K_M is the Michaelis constant.

K_M is often used as a measure of affinity of the substrate for the enzyme. Therefore within the Michaelis-Menton model the lower the K_M the stronger the affinity of substrate for the enzyme. The k_{cat} is a measure of reaction turnover for the enzyme-substrate complex to enzyme and products as shown in equation 1.1.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

$$[k_2 = k_{cat}]$$

The equilibrium approximation: $K_M = \frac{k_{-1}}{k_1}$

The steady-state approximation: $K_M = \frac{(k_{-1} + k_2)}{k_1}$

Equation 1.1 This equation allows the calculation of the affinity of the enzyme for the substrate (K_M). k_2 is the catalytic turnover of enzyme-substrate complex to enzyme and product. The equilibrium approximation is E + S ↔ ES is only slightly disturbed by ES → P, so k_2 is low relative to k_{-1} hence $K_M = k_{-1}/k_1$ whereas the steady-state approximation takes ES → P into account.

1.5.2 Transition state versus ground state

Linus Pauling was pivotal in elucidating the fundamentals of enzymatic catalysis. He proposed that an enzyme binds more tightly to the transition state of a substrate than to the ground state [Pauling, 1948]. Antibodies are molecules that have evolved to bind with high affinity to stable, that is 'ground state', antigens. In 1969 Jencks raised the concept of transition state stabilisation by antibodies. He suggested that an antibody raised to bind the transition state of a reaction could, in theory, catalyse a reaction in a manner analogous to that of the natural enzyme [Jencks, 1969]. This subsequently triggered research into the development of catalytic antibodies.

Initial research into catalytic antibodies met with little success for two main reasons. i) attempts were made to catalyse reactions that had a low background turnover, thus detection of trace amounts of products was difficult. ii) antibodies produced in the 1970's were polyclonal so that the antibodies used would have had multiple specificities towards the antigen. Furthermore, within the polyclonal sera, catalytic antibodies were thought to exist but were only a fraction of total antibody present, consequently turnover levels were poor [Raso and Stollar, 1975; Burd *et al,* 1977; Kohen *et al.*, 1979]. Only the advent of monoclonal antibodies in the mid 1980's allowed the generation of monospecific antibodies in quantities adequate for definitive proof of catalysis. It was in December 1986 that two groups independently demonstrated that monoclonal antibodies directed towards tetrahedral phosphorus species could catalyse the hydrolysis of carboxylate esters and carbonates [Tramantono *et al.,* 1986; Pollock *et al.*, 1986]. One of the antibodies catalysed the hydrolysis of a carbonate by a factor of 10^2 over background. The other antibody catalysed the hydrolysis of a carboxylate ester at a factor of 10^2 over background with both antibodies displaying Michaelis-Menton saturation kinetics. To be true biological catalysts, these antibodies must display the same behaviour of that seen with enzymes.

1.5.3 Catalysis by antibodies

Production of catalytic antibodies has moved beyond demonstrating feasibility of the concept. The main question now is whether catalytic antibodies can ever become efficient and rapid enzymes? Since the first esterolytic antibodies there has been rapid development of catalytic antibodies catalysing a diverse range of chemical reactions; amide hydrolysis [Janda *et al.*, 1988], Claisen rearrangement [Hilvert *et al.*, 1988], Diels-Alder reaction [Hilvert *et al.*, 1989], DNA hydrolysis [Shuster *et al.*, 1992], dis-favoured *exo*-Diels-Alder reaction [Gouveneur *et al.*, 1993] cocaine degradation [Landry *et al.*, 1993] and peptide bond formation [Jacobsen & Schultz, 1994] to name but a few. A unique and potentially useful characteristic of certain catalytic antibodies is their enantiomeric selectivity. These antibodies bind to the D or L form of the substrate, depending on the chiral form of the transition state analogue used to raise the antibody. This stereochemical

selection can be applied to the chiral resolution of a racemic mix of substrate, which potentially might be of use in producing optically pure compounds or chiral drugs, now required by the Food & Drug Administration (FDA) to be demonstrated prior to clinical trials. The catalytic antibody portfolio continues to expand but some recent research efforts have begun to investigate how existing antibody catalysed reactions can be further accelerated. This requires examination of catalytic antibody structures and mutagenic studies in order to elucidate the antibody mechanisms. Unfortunately, relatively few crystal structures of antibodies have been solved to date although the existing structures have provided valuable structural data on their mechanisms of action [Gigant *et al,* 1997; Wedemeyer *et al*, 1997; Charbonnier *et al.,* 1995; Zhou *et al.*, 1994; Haynes *et al,* 1994]. Progress in determining the catalytic mechanisms by mutagenesis has been slowed due to the inconsistency of antibody fragment expression [Kipriyanov *et al*, 1997; Knappik & Pluckthun, 1995; Ulrich *et al.*, 1995].

Catalytic antibodies generally follow Michaelis-Menton saturation kinetics. The values of k_{cat} and K_M and the nature of the interactions present in the combining site have been used to give some insight into antibody substrate complex. A study on polyclonal catalytic antibodies from sheep antisera showed a correlation ($r = 0.87$) between increasing k_{cat} and K_M values, thought to be indicative of destabilisation of antibody-substrate interactions leading to an increase in the rate of catalysis. In addition, a correlation ($r = 0.69$) between an increase of k_{cat}/K_M together with an increasing K_M might indicate stabilisation of the transition state for the reaction [Gallacher, 1993].

Although catalytic antibodies have a relatively slow rate of catalysis when compared with enzymes they are currently the best method for producing a *de novo* protein for a specific catalytic role. Catalytic antibodies are less efficient catalysts than their natural enzymatic counterparts due to enzymes having slowly evolved the ideal environment and mechanism for the processing of its substrate. In addition, catalytic antibodies are more prone to product inhibition. Factors important for catalysis other than transition state stabilisation are also important, such as reactive

groups, oxyanion pockets, enzymatic participation of chemical groups in a reaction, water free microenvironments, destabilisation of the substrate and approximation of reactive groups towards the substrate [Benkovic, 1996; Benkovic *et al.*, 1990; Iverson *et al.*, 1989; Houk, N., 1996].

Antibodies can bind virtually any antigen but antibodies may not have sufficient flexibility in their structure to provide a catalytic site for all reactions. The currently existing abzymes may have already reached their maximum rate of catalysis. This may explain why antibody catalysed reactions are several orders of magnitude lower than those of enzymes catalysing the same reactions. Poor understanding of the antibody mechanisms, poor hapten design, product inhibition and product clearance are all possible obstacles to the efficient and rapid catalysis by antibodies. Early experiments on catalytic antibody selection during screening were based on their strength of binding of the TSA. It is entirely possible that enzymes were not selected by evolution for their affinity for the substrate but upon their activity in catalysing a reaction. This could mean that superior catalytic antibodies are not necessarily those that bind the TSA tightly during the production of catalytic antibodies. In many instances, antibodies that bind their designed TSA have not exhibited any catalytic activity while other antibodies raised against the same TSA have. This may be due to the precise microenvironment of some antibodies being incompatible for catalysis, leading to too tight substrate binding or failure to release product [Haynes *et al.*, 1996]. Therefore the screening method for catalytic antibodies should be directed towards activity rather than binding. This concept has been used to genetically screen for improved Claisen rearrangement antibodies, based on their ability to carry out a key reaction in the aromatic amino acid biosynthesis pathways that is vital for the survival of the mutant strain of *E. coli* used. This bacterium has a mutant gene for metabolite catalysis so that the cell survival is dependant upon successful production of catalytic antibodies that are able to efficiently catalyse the conversion of chorismate to prephenate, the metabolic precursor of phenylalanine and tyrosine [Tang *et al.*, 1991]. However, there are few cases where this genetic screening method has been applied.

1.5.4 Generation of antibodies towards designed haptens

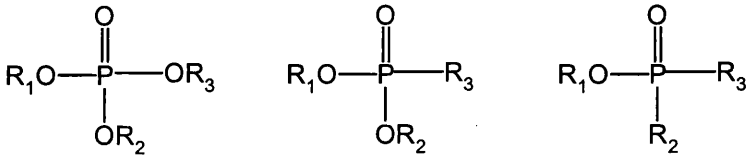
A hapten is a small molecule that can act as an epitope but is incapable of eliciting an antibody response alone. These molecules require conjugation to a carrier molecule to facilite their recognition by the immune system. Haptens, that were transition state analogues of particular reaction transition states, were conjugated to carrier proteins and used to induce catalytic antibodies. As suggested by their name, the transition state structures are unstable or *transient* by nature and so are impossible to use as haptens themselves. However, use of a transition state analogue is not ideal as the conformation of the transition state for most reactions is a 'best guess'. The analogue molecule produced is likely to differ in conformation from that of the true transition state. Therefore the extent of this difference may dictate the success of the approach for particular reactions.

The transition state analogue haptens are conjugated to carrier proteins such as keyhole limpet hemocyanin (KLH) or bovine serum albumin (BSA), necessary to elicit an immune response. Typically, amide bonds are formed between carboxyl groups in the hapten and the ε-amino groups of the surface lysine residues in the carrier proteins [Erlanger, 1980]. The spacer arm between the hapten and carrier is normally at least 6Å in order to prevent steric interference from the carrier during the immunisation process [Nishima *et al.*, 1974].

1.5.5 Antibody esterases

The first antibody catalysts produced were esterolytic antibodies [Tramantano *et al*, 1986; Pollack *et al*, 1986]. Abyzmes that catalysed amide hydrolysis were produced shortly afterwards [Janda *et al.*, 1988]. Most of the haptens designed to catalyse ester/amide hydrolysis mimic the tetrahedral transition state thought to exist during the cleavage of ester and amide bonds. The majority of the haptens used are phosphonates or phosphonamidates from the organophosphorous family (Figure 1.8) which possess a tetrahedral structure through the inclusion of a pentavalent phosphorus atom. Interestingly sulphur based sulphones and sulphonamides have been tried but with no reported success [Thomas, 1994]. An important feature of hapten design is that the presence of aromatic groups within the

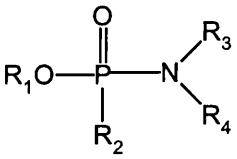
TSA generally leads to more successful catalytic antibody production [Thomas, 1994], probably due to the aromatic moiety eliciting a more vigorous immune response. However, as indicated above, tight binding towards the TSA does not always indicate antibodies with efficient catalysis, highlighting the difficulties in screening for catalytic antibodies.



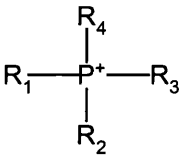
1. Phosphate

2. Phosphonate

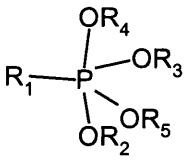
3. Phosphinate



4. Phosphonamidate



5. Phosphonium



6. Phosphorane

Figure 1.8 The nomenclature of organophosphorous compounds on which tetrahedral based transition state analogues are often based.

1.5.6 Elucidating catalytic antibody mechanisms from their crystal structures

An increasingly large array of kinetic data are available on existing and novel catalytic antibodies, while relatively little three-dimensional structural information exists. These data are essential for clear and rapid elucidation of the mechanisms involved in these antibodies [Wilson & Stanfield, 1994] and for identifying the residues involved in covalent or non-covalent interactions, or that simply attribute some catalytic function. At the time of writing, structural data exist on six catalytic antibodies which has allowed the elucidation of their likely mechanisms (Table 1.1). It is important to realise that crystal structures are by nature static representations of the antibody structures and cannot represent their dynamic state. Thus, it is necessary when examining crystal structures to take account of the importance of flexibility in catalysis, as with enzymes.

CNJ206

The antibody CNJ206 possess esterolytic activity towards a *p*-nitrophenyl ester displaying a k_{cat}/k_{uncat} of 1.6×10^6. It has a long but shallow groove at the interface of the heavy and light variable chains. The residues thought to be involved are Tyr^H50, Tyr^L96 and His^H35 which hydrogen bond the phosphoryl moiety of the TSA while the Asp^L55 binds the $-NO_2$ group. Hydrogen bonds are thought to form between the hapten and the Tyr^L91, Tyr^L49, Tyr^L96 Tyr^H50 and Tyr^H59 residues. Binding of hapten (Figure 1.9) or substrate in this antibody is thought to produce a dramatic shift in 11 residues of the H3 loop triggering a conformational change from a shallow groove to a deep pocket. [Golinelli-Pimpaneau *et al.*, 1994; Charbonnier *et al.*, 1995].

1F7

The 1F7 antibody displays chorismate mutase activity with a k_{cat}/k_{uncat} of 2.5×10^2. It was raised towards a molecule that inhibits chorismate mutase that is found in bacteria, fungi and plants (Figure 1.9). The chorismate mutase enzyme catalyses a key step in the biosynthesis of phenylalanine and tyrosine. This allowed the antibody to act as an *in vivo* catalyst for this metabolically essential reaction. The cells were auxotrophic mutants and so were deficient in the enzyme used for producing phenylalanine and tyrosine, were only able to survive if expressing the antibody [Tang *et al*, 1991]. It is thought that the antibody stabilises the same conformationally restricted pericyclic transition state as that occurring in the uncatalysed reaction. The structure of antibody 1F7 in complex with its hapten has been solved. The structural data allow comparison of the active site of the enzyme with that of the antibody, potentially providing evidence of any similarity between the two catalytic mechanisms. The base of the binding site appears relatively hydrophobic; Asn^H35, Trp^H47, Ala^H93, Phe^H100b, Leu^L96 and the walls consist of polar residues; Asn^H33, Asn^H50, Asn^H58, Arg^H95, Tyr^H100 and Tyr^L94. Hydrogen bonds are thought to form between Asn^H33 and the hydroxy group of the hapten and between the side chain hydroxy group of the Tyr^L94 and the 3° carboxyl group of the hapten. Water molecules may form hydrogen bonds between the haptens 2° carboxyl group and the residues Arg^H95, Asp^H101. However when compared with

the original enzyme, the antibody has rotational limits upon some residues and in addition, fewer electrostatic interactions appear to be used to stabilise charge separation during catalysis. Hapten recognition is dominated by van der Waals contacts but many residues probably make contact with the hapten through a limited number of hydrogen bonds [Haynes *et al.*, 1994].

17E8

The structure of an esterolytic catalytic antibody 17E8 bound to its hapten (Figure 1.9) was solved at 2.5 Å resolution [Zhou *et al.*, 1994]. This antibody catalyses the hydrolysis of nor-leucine or L-methionine phenyl esters, with the structural and kinetic data supporting a hydrolytic mechanism similar to that seen in serine proteases. The antibody's combining site has a Ser-His diad proximal to the phosphorous atom of the bound hapten which resembles the His-Ser-Asp triad that is seen in serine proteases. In addition the site contains a Lysine residue thought to stabilise oxyanion formation and a binding pocket which functions in substrate recognition of the norleucine/methionine side chains. Fourteen of the antibody's residues are involved in binding the hapten with the phenyl ring bound in a pocket formed by the side chains of $Tyr^{L}36$, $Leu^{L}89$, $Arg^{L}96$, $Phe^{L}98$, $Val^{H}37$, and $Trp^{H}47$ residues. The side chain of the hapten is contacted by $Gly^{L}34$, $Leu^{L}89$ and $Tyr^{L}91$ and its phosphonate group has electrostatic interactions with $Lys^{H}93$ and $Arg^{L}96$ residues. This mechanism suggests a convergence to the serine proteases family of catalytic structures which have evolved in some enzymes. *Atomic contacts were seen between the $Ser^{H}95$ and $His^{H}35$ towards the phosphorous atom of the TSA.* During catalysis the pH is likely to be important as at a neutral pH, $Ser^{H}99$ accepts a hydrogen bond from the donor $Tyr^{H}97$. However at pH 9.5, this hydrogen bond is removed and the serine residue is thought to rotate allowing nucleophilic attack upon the substrate carbonyl atom. The $His^{H}35$ residue accepts a proton from the freely rotating $Ser^{H}99$ residue, which allows it to form the covalent acyl intermediate with the substrate. The $Lys^{H}97$ stabilises the subsequent charge which builds upon the carbonyl atom. The phage display technology described earlier has been used to optimise binding of this catalytic antibody to its TSA resulting in identification of variants with improved catalytic efficiencies [Baca *et al.*, 1997].

48G7

The crystal structure of the esterase catalytic antibody 48G7 was solved with both the antibody complexed and uncomplexed to the hapten (Figure 1.9)[Wedemayer *et al.*, 1997]. The antibody catalyses the hydrolysis of *p*-nitro phenyl esters and carbonates. Examination of the active site with and without the TSA displayed no significant changes due to hapten binding and has been described as "mutual fit" where antigen binding causes relatively minor structural changes [Webster *et al.*, 1994]. The hapten fits into a pre-formed pocket of approximately 10Å in depth and shows no involvement of water molecules within the active site. The structure did illustrate an emerging theme seen with esterolytic antibodies where binding phenyl haptens (& nitro-phenyl haptens). The aryl end is buried deep into the binding pocket while the phosphonate and phenyl moieties are responsible for the majority of the antibody-hapten interactions. The complex shows that the aliphatic linker and carboxylic acid group is surrounded by six tyrosine residues; $^{L}32$, $^{L}91$, $^{L}94$, $^{L}33$, $^{H}95$, $^{H}96$. The phenyl ring is contacted by Tyr$^{L}91$ and Tyr$^{H}99$, the phosphorus atom of the hapten makes no direct contact with the antibody while its oxygen atoms form hydrogen bonds with Arg$^{L}96$, Tyr$^{H}33$ and van der Waals interactions are seen with His$^{H}35$. Six tyrosine residues are in the active site. There has been speculation that antibodies containing high numbers of aromatic residues within the active sites can lead to π-π orbital stacking with the phenolic moieties of the hapten restricting product release and possibly causing product inhibition [Thomas, 1993]. However strict π-π stacking was not observed between the phenyl moiety and aromatic residues within the 48G7 antibody.

D2.3

Crystal structures have been solved of the esterolytic antibody D2.3, complexed with a substrate analogue, transition state analogue (Figure 1.9) and with one of its reaction products, as well as in the unliganded form. The antibody contains an narrow groove in its active site but provides a deep pocket for the substrate to bind. A negatively charged oxyanion intermediate is formed through nucleophilic attack by water molecules. A tyrosine is thought to be integral in activating the ester

substrate after which it helps stabilise the oxyanion intermediate with the help of a proximal asparagine residue [Gigant *et al.*, 1997].

NPN43C9 Fv

Crystal structures are not the only structure-based method of elucidating catalytic antibody mechanisms. Modelling can be a rapid alternative but is a method that must be viewed with substantially less confidence. A model of the antibody NPN43C9 Fv was built and used to select residues for mutagenesis studies that were thought to be important in catalysis [Stewart *et al*, 1994]. The $Arg^{L}96$ side chain at the base of the combining site was thought to interact with the phosphonamidate of the hapten (Figure 1.9). When mutated all catalytic activity was eliminated. Either residue $His^{L}91$ or $His^{H}35$ was thought to be a nucleophile involved in the formation of a covalent intermediate during catalysis.

The dielectric character of an enzyme or catalytic antibody's catalytic microenvironment is also very important. Ionisation of catalytic residues is dependent upon their pKa, and can be an important, if not critical factor in determining the optimum pH for the catalytic reaction. Often the pH at which the reaction is to be performed will be lower than the optimum value for catalytic turnover. As these pH extremes can have a destructive effect on the structure of the antibody, a balance must be achieved.

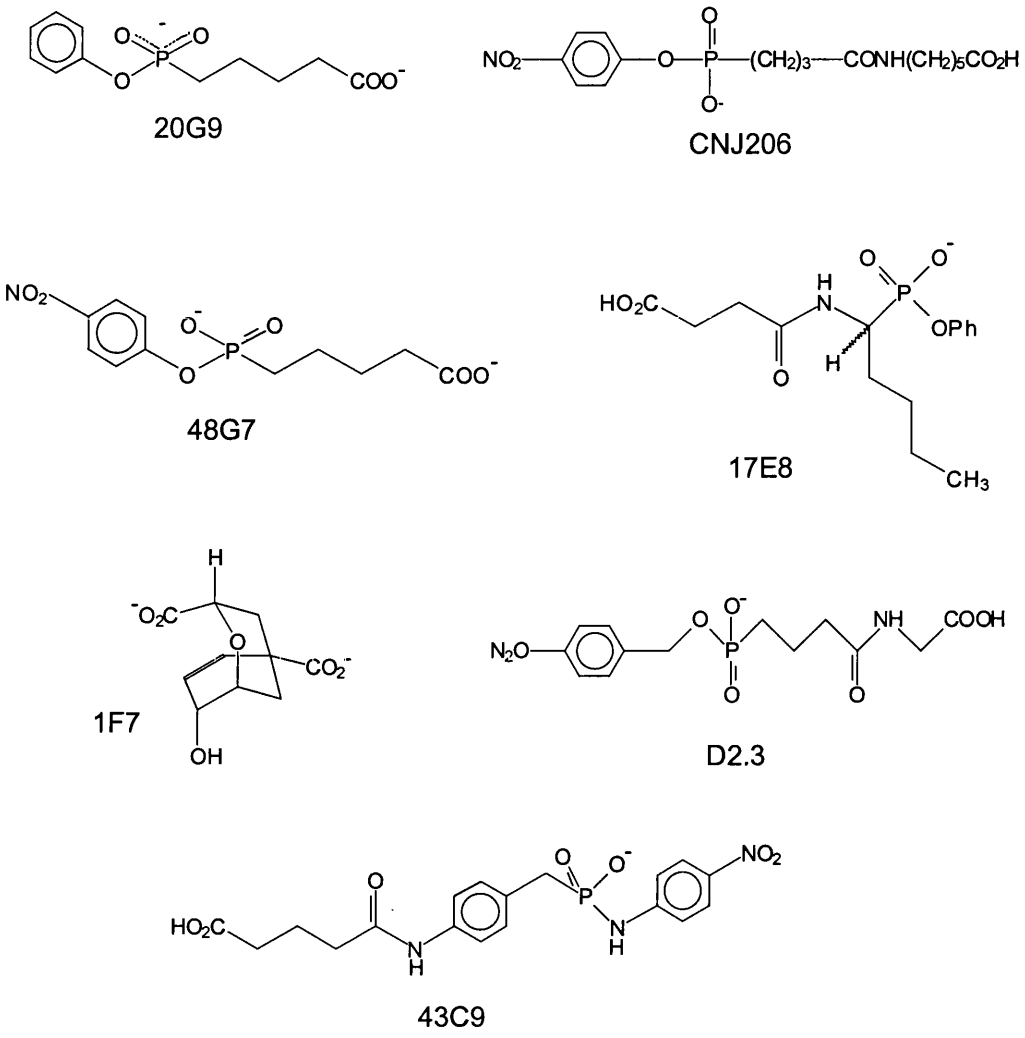


Figure 1.9 The transition state analogue haptens used to raise the catalytic antibodies: 20G9, CNJ206, 48G7, 17E8, 1F7, D2.3, 43C9. All, excluding 1F7, included a tetrahedral phosphonate group that mimicks the tetrahedral shape of the reaction intermediates formed.

Table 1.1 Residues observed in crystal structures and models of catalytic antibodies which are involved in the antibody's catalytic mechanism. Residues numbers that coincide have been notated in bold *Non-esterolytic antibody

Abzymes	*Light chain residues*	*Heavy chain residues*
CNJ206	$Y^{L}49$, $D^{L}55$, **$Y^{L}91$, $Y^{L}96$**	**$H^{H}35$**, $Y^{H}50$, $Y^{H}59$
17E8	**$G^{L}34$**, $Y^{L}36$, $L^{L}89$, **$Y^{L}91$, $R^{L}96$**, F^{L}	$V^{H}37$, **$W^{H}47$**, $K^{H}93$
48G7	$Y^{L}32$, $Y^{L}33$, **$Y^{L}91$, $Y^{L}94$, $R^{L}96$**	**$H^{H}35$**, $L^{H}95$, $Y^{H}96$, $Y^{H}99$
D2.3	**$A^{L}34$, $G^{L}91$, $Y^{L}96$**	$Y^{H}100d$, **$W^{H}47$, $H^{H}35$**
NPN43C9 Fv	**$R^{L}96$, $H^{L}91$**	**$H^{H}35$**
1F7*	**$Y^{L}94$, $L^{L}96$**	$N^{H}33$, **$N^{H}35$**, **$W^{H}47$**, $N^{H}50$, $R^{H}58$, $A^{H}93$, $F^{H}100b$

1.5.7 The esterolytic antibody, 20G9

The antibody 20G9 was generated in 1988 against the transition state analogue phenylphosphonate, **1** [Dufor *et al.*, 1988] and catalyses the hydrolysis of phenylacetate. The TSA was coupled to the carrier protein keyhole limpet haemocyanin (KLH) and bovine serum albumin (BSA) at a ratio of 15:1 (hapten/carrier). The KLH-phosphonate conjugate, buffered in 10mM phosphate and 150mM NaCl pH7.4, was mixed with Freunds complete adjuvant and used to immunise BALB/c mice over a period of months. The mouse was later sacrificed and its spleen cells fused with Sp2/o myeloma cells to form cell hybridomas, from which thirteen clones were successfully isolated which bound the phenylphosphonate TSA. The antibody's substrate, phenyl acetate, spontaneously hydrolyses with a k_{uncat} of 4.9×10^{-4} min^{-1}, with the suggested mechanism displayed in Figure 1.9.

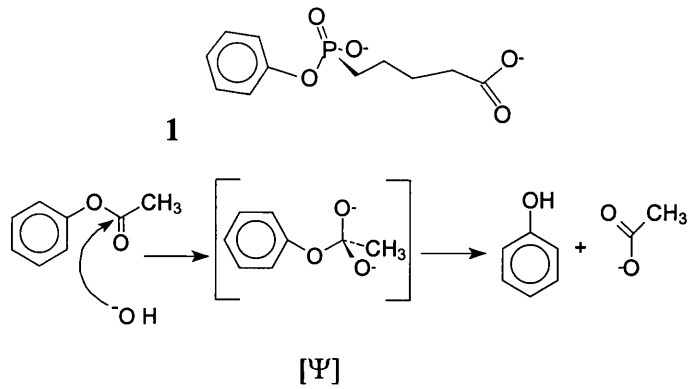


Figure 1.10 The phenylphosphonate hapten (**1**) to which the 20G9 antibody was generated. The possible mechanism of the hydrolysis of phenyl acetate to phenol and acetic acid via a tetrahedral transition state [Ψ].

The 20G9 monoclonal antibody was purified by Protein A affinity chromatography and homogeneity confirmed by SDS-PAGE. The antibody was partly characterised in 1990 where its possible use as a biosensor was investigated [Blackburn *et al*, 1990] but the majority of kinetic characterisation was performed by Martin *et al* in 1991. This characterisation discovered two discrete kinetic rates; a pre-steady state burst and a steady-state phase thought to exist possibly due to product inhibition. The kinetic parameters were determined for the initial and steady-state velocities: pre-steady state k_{cat} = 4.9±(0.5) min^{-1}, K_M = 300±(70) μM and a steady state k_{cat} = 4.9±(0.5) min^{-1}, K_M = 30 ±(70) μM. The pH profile for the steady state displayed an acid limb pK of 9.6. Tetranitromethane (TNM) was used to selectively nitrate active tyrosines and resulted in an almost complete loss of activity. This indicated tyrosine involvement in the catalytic site although their specific function(s) and whether this indicated the possible formation of an acyltyrosyl intermediate can only be a subject of speculation. It may have been that product binding of phenol to this intermediate was responsible for the inhibition observed.

1.5.8 A future for catalytic antibodies?

The opportunities for the application of catalytic antibodies within industry appear limited and it is difficult to envisage that they will be utilised within large industrial processes even if the rate could be improved. However a niche may exist at a clinical level with the activation of pro-drugs of *de novo* design or in destruction of bacterial or viral coat proteins. At an analytical level they have the potential to be used as novel restriction proteases using their ability as amidases. Catalytic antibodies have been investigated as biosensors [Blackburn *et al*, 1990] and so a possible future for abzymes may exist in micro-analytical systems with a role as a robust biological component when an enzyme does not exist. The additional benefits are due to the similar structures and properties of catalytic antibodies, which can allow standard immobilisation, stabilisation, calibration, storage and labelling procedures thus avoiding lengthy optimisation studies that would be necessary for different enzymes. In addition, antibodies are often found to be more

robust than enzymes, adding flexibility with regard to reaction temperature and pH as well as improved short term storage.

1.6 THE AIM OF THIS THESIS

The aim of this thesis is to kinetically characterise the catalytic scFv 20G9. As described, the 20G9 antibody has previously undergone some biochemical and kinetic characterisation, discovering the burst and steady-state phases of the reaction. It has been speculated that tyrosine residues have a direct involvement in the catalytic mechanism of this antibody. This thesis aims to reinvestigate this and identify those residues involved in the catalysis of this antibody.

Molecular modelling will be used to identify residues that have a potential involvement in the mechanism. This will be carried out using *AbM*, an antibody modelling package developed in this laboratory, which allows the production of reasonably accurate three-dimensional structures of the antibody and its binding site. Residues selected as candidates for involvement within the binding site will primarily be selected based upon their proximity to the antibody binding site. The number of residues will then be narrowed down using the existing kinetic evidence available on 20G9 and by reference to the involvement of specific residues within other esterolytic antibodies. The selected residues will then be mutated in two stages: primary and secondary candidates for mutation, based upon their relative likelihood to be involved in the catalysis.

The residues selected for mutation will be mutated using PCR site-directed mutagenesis. Due to the low error rates seen with some polymerases during PCR, this is considered an easy and rapid method with which to mutate the selected residues. The mutated section can then be easily re-cloned into the original vector, or into another suitable vector for high-level expression. All mutants will be fully sequenced to check for polymerase errors before expression of the mutated antibody. The simplest way of identifying successfully cloned mutants is to express

the antibodies using an affinity tag. This will facilitate purification of the bacterial expressed antibody fragment from the cell media.

The mutants will then be characterised, relative to the wild-type scFv and previous kinetic data from the 20G9 antibody, for kinetic activity. Changes in the steady-state and burst rates will be examined taking careful note of any changes to the K_M and k_{cat} of the steady-state and kinetic burst. Observed changes will be correlated to mutations introduced into the antibody, allowing greater understanding of the kinetic mechanism of this catalytic antibody.

CHAPTER 2

2 STANDARD MATERIALS AND METHODS

This chapter describes the general biochemical and molecular biology techniques used in the experiments described in this thesis. Specific methods used for particular experiments are detailed within that particular chapter.

2.1 MATERIALS

2.1.1 Bacterial strains used for cloning:

Table 2.1 *E. coli* strains used for cloning

Strains	Source	Genotype	Reference
XL-1 Blue	Stratagene	[F'::Tn10 (Tc^R) $proA^+$ $proB^+$ $lacI^q$ lacZΔM15] recA1 endA1 gyrA96 (NaI^R) thi hsdR17 ($r_k m_k^+$) supE44 relA lac	Bullock, W.O. *et al.* (1987).
DH5α	Stratagene	endA1 hsdR17 ($r_k m_k^+$) supE44 thi^{-1} recA1 gyrA (NaI^R relA1 Δ(lacZYA-argF)U169 (ϕ80lacZΔM15)	Hanahan, D. (1983)
JM109	Promega	thiI rpsL $endA^+$ sbcB15 hsdR4 supE D (lac-proAB)/F' [traD36, $proAB^+$ $lacI^q$ lacZΔM15]	Yanisch-Perron, C. *et al.* (1985)

2.1.2 Bacterial strains used for protein expression:

Table 2.2 *E. coli* strains used for protein expression

Strains	Source	Genotype	Reference
TOPP2	Stratagene	Rif^r [F' $proAB^+$, Tn10 (Tet^r)]	Hatt, J., *et al* (1992).
TG1	G. Winter MRC	sue hsd ΔS thiS[lac-proAB] F' [traD36 $proAB^+$ $lacI^q$ lacZ ΔM15	Gibson, T. (1984)

2.1.3 Enzymes, chemicals and oligonucleotides

All enzymes were obtained from New England Biolabs (UK) Ltd., Hitchin, UK and chemicals from Sigma-Aldrich Chemical Co. Ltd., Poole, UK, Fisons Ltd., Loughborough, UK and BDH Ltd., Poole, UK unless stated otherwise. All oligonucleotides used were custom made by Perkin-Elmer Ltd., Warrington, UK.

All were HPLC purified and supplied lyophilised. Throughout, the water was Millipore Q-Plus filtered water, Millipore, Watford, UK.

2.1.4 Bacterial cell culture

All cell culture materials were obtained from Difco Laboratories, Michigan, USA.

LB	(1 Litre) 10g Tryptone, 5g Yeast Extract, 5g NaCl
2xYT	(1 Litre) 16g Tryptone, 10g Yeast Extract, 10g NaCl
XL	(1 Litre) 16g Tryptone, 15g Yeast Extract, 10g NaCl

2.2 BIOCHEMICAL AND DNA MODIFICATION TECHNIQUES

2.2.1 Agarose gel electrophoresis

Nearly all sizes of DNA can be resolved using agarose gel electrophoresis by reducing or increasing the amount of agarose. The DNA within the gel was stained with ethidium bromide and visualised under long or short wave ultraviolet (UV) light. Gels were prepared by dissolving 1% w/v agarose (Gibco) in TBE buffer (89mM Tris-borate, 89mM boric acid and 2mM EDTA) and boiling in a microwave oven. The gel was allowed to cool to approximately 50°C at which point ethidium bromide was added to a concentration of 0.5μg/ml. The molten agarose was poured onto a horizontal gel bed that contained spacers and a well-forming comb. Once set the comb and spacers were removed and the gel submerged in TBE buffer. DNA samples were taken up into loading buffer (0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol, 40% w/v sucrose) at a ratio of 1:5, dye:sample. Samples, 100bp ladder and 1Kb ladder molecular weight markers (Gibco) were loaded onto the gel and electrophoresed at 80V until the marker dyes had migrated sufficiently. DNA was visualised by fluorescence using short-wavelength ultraviolet light (UV). If the DNA was to be extracted then the above procedure was identical except that low melting point (LMP) agarose (Gibco) was used and the procedure performed at 4°C to prevent the gel melting. Damage to the DNA was prevented by brief visualisation with long-wave UV light.

2.2.2 Concentrating DNA by ethanol precipitation.

3M sodium acetate (CH_3CO_2Na) pH 5.5 was added to the DNA to a concentration of 300mM. 2.5 volumes of absolute ethanol were added and the solution gently mixed and incubated overnight at -20°C. The precipitated DNA was centrifuged at 15,000 x g for 30 minutes at 4°C and the supernatant carefully removed by aspiration. The DNA was dried using a SpeediVac vacuum centrifuge at 500 x g for 30 minutes and resuspended in 5-10µl sterile water or TE buffer and stored at -20°C.

2.2.3 Determination of DNA concentration

5µl of the DNA solution was diluted to 1ml with water. Spectrophotometer readings were taken at A_{260} and A_{280}. The A_{260} reading was measured and used to estimate the concentration of the DNA using the following equation:

1 A_{260} = 50 µg/ml double stranded DNA

1 A_{260} = 33 µg/ml single stranded DNA

The ratio (A_{260}/A_{280}) provided an estimation for the purity of the DNA preparation. A ratio of 1.8 indicated high purity. A ratio greater than 1.8 suggested RNA contamination while a lower figure indicated protein or solvent contamination, i.e. phenol.

2.2.4 Restriction enzyme digestion

Restriction digests using NEB enzymes were performed using 100-300ng DNA in a volume of 60µl. All digests were performed in the appropriate NEB enzyme buffer (NEBuffer Activity Chart for Restriction Endonucleases, NEB catalogue, 96/97). Some NEB enzymes required the addition of BSA to a concentration of 100µg/ml. Digestion of plasmid DNA required enzyme at 1 unit per 1 mg DNA whereas digestion of PCR extension products required 3 units per 1 mg DNA. Digestions were generally performed at 37°C overnight and inactivated by heating at 65-80°C for 15-30 minutes.

2.2.5 Ligation of sticky and blunt ended DNA

A Fast Link ligation kit (Epicentre Technologies, Wisconsin, USA) was used to perform rapid ligations on sticky and blunt ended fragments. Approximately 25ng of digested PCR insert and 10-15ng of digested plasmid DNA were added together giving a ratio of 2:1 insert:plasmid. For blunt ended ligations the ratio was increased to 10:1 to achieve the greatest number of recombinants. 1.5µl of 10x ligation buffer (100mM NaCl, 0.1mM EDTA, 0.1% v/v Triton X-100, 1mM dithiothreitol (DTT), 50mM Tris-HCl pH 7.5) was mixed with 1.5µl of 10mM ATP. 1.0 µl of T4 DNA Ligase (2 U/µl) was added to the DNA. The reaction volume was made up to 15µl using sterile H_20. The tubes were allowed to incubate at room temperature for at least 1 hour after which the reaction was stopped by heating at 70°C for 15 minutes to inactivate the DNA ligase (Michelsen, B.K., 1995). 10% of the reaction mix was used directly to transform competent *E. coli* cells.

2.2.6 Spin-column chromatography

Spin-column chromatography was used for the removal of primers, restricted end fragments (<10bp), desalting and buffer exchange from small volumes (< 50µl).

MicroSpin columns (Pharmacia Biotech Inc., St. Albans, UK) are packed with Sepharyl resin of various pore sizes (200,300,400) and are pre-equilibrated with TE buffer (pH 7.6).

The column resin is resuspended by shaking, the cap is loosened and the sealed base snapped off. The column was placed into a 1.5ml Eppendorf tube and centrifuged in a microfuge at 3,000 x g for 1 minute. If appropriate, 0.5ml of a buffer was placed into the column and the above spin step repeated. The column was moved into a fresh Eppendorf tube and the DNA solution carefully added onto the centre of the resin. The column was centrifuged at 3,000 x g for 2 minutes and the flowthrough collected.

2.2.7 DNA recovery from a low melting point (LMP) agarose gel

DNA was visualised using long wave UV light and the band carefully cut out using a clean razor blade. The gel fragment containing the DNA was weighed to

determine its approximate volume and melted at 65°C. 10x Agarose buffer (NEB) was diluted to 1x concentration in the liquefied agarose, mixed and allowed to cool to 40°C. 1 µl of β-agarose enzyme ((1 U/ µl) NEB) was added per 100µl 1% w/v agarose gel followed by incubation at 40°C for 2 hours. 3M Sodium acetate, pH 5.5 was added to a final concentration 300mM and the mixture chilled on ice for 15 minutes. Any undigested agarose was removed by centrifugation at 15,000 x g and the supernatant volume measured. 2 volumes of iso-propanol (pre-cooled at -20°C) were added and the mixture incubated at -20°C for at least 30 minutes. Following the precipitation the solution was centrifuged at 15,000 x g at 4°C for 15 minutes. The alcohol was carefully removed by aspiration leaving approximately 50µl solution to which 1ml of 70% ethanol was added. The tube was inverted 2-3 times and centrifuged at 15,000 x g for 5 minutes. The supernatant was removed by aspiration and the remaining alcohol evaporated with a short spin in a SpeediVac vacuum centrifuge. The pelleted DNA was resuspended in 10-15 µl of Tris-HCl pH 7.5 (TE) or sterile water.

2.2.8 Polymerase chain reaction (PCR) using Vent DNA polymerase

For high fidelity primer extensions, Vent DNA polymerase (NEB) was used which utilises its 3'-5' exonuclease activity for proof-reading of the extension product. 5µl of a 10x stock of dNTPs (dATP, dCTP, dGTP, dTTP at 3mM) (Pharmacia Biotech), 5µl of 10x Vent polymerase buffer (100mM KCl, 200mM Tris-HCl pH 8.8, 100mM $(NH_4)_2SO_4$, 0.1% Triton X-100) and 0.5 µl of BSA (10mg/ml) were added to 0.2ml thin-walled tubes. The tubes were placed under short-wave UV light for 5 minutes to eradicate any contaminating DNA. 1 µl of each oligonucleotide primer (50pmoles/µl), 10ng of template DNA, 0.5µl of Vent DNA polymerase (2 U/µl) were added and the volume made up to 50µl using sterile Milli-Q water. A drop of mineral oil was added to the tube to cover the reaction and prevent evaporation. All PCR's performed included a negative control (no DNA template added) and a positive control on an existing template known to give PCR products, allowing testing of the various parameters within the PCR such as buffers, enzyme and the temperature cycle.

The thermo-cycler, model PTC-100 (Perkin-Elmer, USA) was programmed to heat to 94°C for 2 minutes to separate any hybridised DNA followed by cooling to the annealing temperature which was held for 2 minutes. For complementary oligonucleotides of 11-20 bases, the annealing temperature was calculated using the Wallace equation: $T_m = 4(G + C) + 2(A + T)$ (Suggs *et al.*, 1981). The calculated annealing temperature allowed optimal binding of the oligonucleotide primer to the template. This was followed by heating to an extension temperature of 72°C, allowing one minute per kilobase of expected extension product (excessive extension time lead to degradation of the extension product due to the polymerase's exonuclease activity). The reaction was then heated to the denaturing temperature of 94°C for 2 minutes leading to separation of DNA (primers, template, products). The reaction was cycled back between the annealing, extension and denaturing steps 28 times. The final step involved cooling to 72°C and maintaining this temperature for 1-2 minutes, which ensured that all extension products were double, stranded. Taq DNA polymerase was used mainly for screening exercises as the PCR products were not used for any cloning experiments due to risk of polymerase misincorporations. DNA polymerase fidelity is described in chapter 4.

2.2.9 Design of oligonucleotides

Oligonucleotide primers were designed in two steps. The first was to decide the function of primer (sequencing primer, PCR primer, PCR primer with base mismatches). The site of template binding was based primarily upon its proposed function (i.e. sequencing primer, introduction of restriction sites) but the following factors were also taken into consideration:

i) no runs of a single base (no more than 4) especially G/C rich domains.

ii) primers were at least 18 bases long to ensure good hybridisation and raise the stringency of the PCR.

iii) melting temperature of primer > 45°C.

The second step involved computer assessment of the designed primer using publicly available software (Amplify, Bill Engels 1992. University of Wisconsin). This allowed the primer to be checked for dimerisation against itself, partner primers and predict non-specific binding. This allows avoidance of 'primer-dimers'

where the oligonucleotide dimerise to itself or its partner primers reducing the amount of available primer within the reaction. The primer software also allows prediction of the PCR product size and highlights particular areas of non-specific binding providing the sequence of the DNA template is available.

If a restriction site was necessary within a desired DNA product, the section "Cleavage close to the end of DNA fragments (oligonucleotides)", the reference appendix of the 96/97 NEB catalogue, was consulted to ensure that oligonucleotides were designed to produce PCR fragments that would restrict satisfactorily.

2.2.10 Automated DNA sequencing - Fluorescent DyeDeoxy terminator method

For automated sequencing, 300-500ng of double-stranded DNA or 30-90ng of PCR product DNA was required. Any DNA sequenced was routinely purified by extraction from LMP agarose gel. 3.2 pmol of HPLC purified primer was added to a 0.2ml thin-walled tube containing the DNA and the volume made up to 12 µl with Milli-Q water. The samples were sequenced on a 377 DNA Sequencer ABI Prism (Perkin-Elmer) using the fluorescent dye terminator method at the *Automated DNA Sequencing Service*, School of Biology and Biochemistry, Bath University. The primer concentration was determined using the following formula:

$$\frac{pmol}{\mu l} = \frac{100(A_{260})}{(1.54nA + 0.75nC + 1.17nG + 0.92nT)}$$

[n = number of A, C, G or T]

All DNA sequences were checked against the sequence chromatograph to ensure that each base had been correctly identified and 'called' by the sequencing software.

2.2.11 Manual DNA sequencing - Sanger Coulson Protocol

Manual DNA sequencing [Sanger at al., 1977] was performed using the Sequenase kit v.2.0 (US Biochemicals, Cleveland, USA) based upon the manufacturer's modified protocol. The two sequencing plates were soaked in 5M KOH overnight, washed with water and polished with absolute ethanol. 1.2ml of 10% v/v acetic acid, 25ml of absolute ethanol, 200µl of silane A-174 adhesion promoter were

mixed and carefully poured over the backplate. After 5 minutes the backplate was washed with water and polished with absolute ethanol.
Using a tissue, 5ml of dimethydichlorosilane solution was carefully wiped onto the surface of the notched sequencing plate. The plate was allowed to dry for 5 minutes, rinsed off with distilled water and dried. Spacers were placed between the edges of the two plates and the plates taped together with extra tape at the bottom of the plates to ensure a good seal.

The gel matrix was prepared by mixing 25g Urea, 15ml Milli-Q water, 10ml 40% *bis*-acrylamide (AccuGel) and 5ml 10x TBE (890mM Tris-borate, 890mM Boric acid, 20mM EDTA). The mixture was dissolved by warming at 37°C and 300µl ammonium persulphate (APS) (100mg/ml) added. 30µl N,N,N',N',-tetramethylethylene-diamine (TEMED) was added, the solution mixed and poured into the prepared plates. The reverse edge of the sequencing comb was placed into the notched section of the plates and bulldog clips applied at various positions over the plates to apply pressure to the setting gel.

1µl of 5M NaOH was added to 5µg of double stranded DNA to separate the DNA and the volume made up to 15µl with water. After 5 minutes incubation at room temperature the DNA was passed through a MicroSpin S-200 column (Pharmacia) and the recovered DNA temporarily stored on ice to prevent re-annealing. 7µl of the DNA solution was mixed with 2µl of Reaction buffer and 1µl primer (10pmol/ µl). This template mixture was heated to 65°C for 5 minutes and allowed to cool to room temperature for 30 minutes, then stored on ice for up to 4 hours. Four 1.5ml Eppendorf tubes were labelled A, T, G, C. 2.5 µl of the appropriate ddATP, ddTTP, ddGTP, ddCTP were dispensed into the tubes respectively and the tubes incubated at 37°C until required. 1 µl of labelling mix was added to 4µl Milli-Q water and stored on ice. 0.5µl of DNA polymerase was mixed with 3.5µl of enzyme dilution buffer.

1µl DTT, 2µl of diluted labelling mix, 0.5µl α-^{35}S dATP (1200µCi) and 2µl DNA polymerase were added to the template mixture and incubated at room temperature for 5 minutes. 3.5µl of this mixture was added to each of the labelled tubes and incubated at 37°C for 5 minutes after which 4 µl of Stop solution was added. The tubes were incubated at 75°C for 5 minutes before loading onto the gel.

The tape at the base of the gel was slit, the comb removed and the plate fixed into the sequencing apparatus. The top and bottom of the plates were immersed in TBE and the gel pre-warmed at 45 watts, 1700 volts for 30', purging the gel occasionally to remove excess urea build-up. The toothed edge of the comb was carefully inserted 2mm into the gel to form the wells and sequencing samples loaded in A, C, G, T order. The gel was run at 45 watts, 1700 volts for 4 hours or until the lower sequencing dye neared the base of the gel.

The gel was removed from the sequencing apparatus and the tape removed. The plates were carefully prised apart and the sequencing gel attached to the back plate was soaked in 1 litre 10% v/v methanol, 10% v/v acetic acid for 5 minutes. The gel was dried onto the plate in an 80°C oven for 30 minutes after which the radio-label was measured using a Geiger counter to verify that the procedure had worked. In a darkroom, autoradiography film was applied to the dried gel and allowed approximately 48 hours to expose. Finally film was processed in an automatic autoradiography processor and the sequence read.

2.2.12 Radioactive labelling of DNA

PCR extension products were radiolabelled to observe restriction enzyme efficacy by the addition of [γ-^{32}P] to their 5' termini. 10 pmoles of DNA was mixed with 2µl 10x polynucleotide kinase buffer (700mM Tris-HCl pH 7.6, 100mM $MgCl_2$, 50mM DTT, 20 pmoles of γ-^{32}P (>5000Ci/mole), 1 unit of T4 polynucleotide kinase (10 U/µl) and made up to a total volume 20µl with Milli-Q water. The mixture was heated to 70°C for 5 minutes to separate strands, briefly chilled on ice and incubated at 37°C for 30 minutes. The reaction was stopped by inactivation of the enzyme at

65°C for 20 minutes. Incorporation of radiolabel was approximately 60% that gave the labelled DNA an activity of between 10^7-10^8 cpm/μg.

2.2.13 SDS-PAGE Protein electrophoresis

Proteins were analysed by electrophoresis through a sodium dodecyl sulphate polyacrylamide matrix (SDS-PAGE). A 12% w/v resolving gel was prepared by mixing 4ml 30% w/v bis-acrylamide solution (ProtoGel) with 3.40ml H_20 and 1.5M Tris-HCl pH8.8. 100μl of 10% w/v SDS, 40μl of 25% w/v ammonium peroxodisulphate (AMPS) and 4ul of TEMED were mixed. 3.5ml of the mixture was poured into the vertical plates of the Mini-Protean II system (Bio-Rad, Herts., UK). 300μl of water was carefully added to ensure a smooth interface between the stacking and resolving gel and the gel allowed to set. A 6% w/v stacking gel was prepared by mixing 1.7ml 30% w/v *bis*-acrylamide solution with 6.7ml H_20, 1.25ml 1M Tris pH6.8, 100μl 10% w/v SDS and 40μl 25% w/v AMPS. 10μl of TEMED was added and the mixture poured onto the resolving gel, a comb inserted and the gel allowed to set. Samples were mixed with loading solution (50mM Tris-HCl, pH6.8, 25% v/v glycerol, 0.1% w/v bromophenol blue, 6% w/v SDS, 4% v/v β-mercaptoethanol) at a ratio of at 5:1 (sample:loading solution). Electrophoresis was performed in tank buffer (250mM Tris-HCl pH 8.3, 192mM glycine, 0.1% w/v SDS) at 200V until the bromophenol dye passed to the bottom of the gel. Proteins were stained by submerging the gel in Coomassie dye solution (0.1% w/v Coomassie Brilliant Blue R250, 50% v/v methanol, 10% v/v acetic acid) for 2-3 hours and destained in destain solution (10% v/v methanol, 10% v/v acetic acid) for 1-2 hours. A silver staining kit (Bio-Rad) was used when high sensitivity was required, (lower detection 0.1ng/mm^2), following the manufacturers instructions. Stained gels were wrapped in cellulose film (BIO-RAD) and dried under vacuum at 60°C.

2.2.14 HRP Western blotting

Six pieces of 3MM filter paper (Whatmann) and one piece of Nitrocellulose Hybond-C membrane (Amersham, Bucks., UK) were carefully cut to the size of the

mini-gel and soaked in 100ml transfer buffer (39mM glycine, 48mM Tris-HCl pH 8.3, 0.037% w/v SDS, 20% v/v methanol) for 10 minutes. Three pieces of the filter paper were placed onto the blotting apparatus, 2117 Multiphore II Electrophoresis unit (LKB Bromma). The nitro-cellulose paper was placed on top of the blotting paper and the mini-gel onto the nitro-cellulose paper. The remaining pieces of the blotting paper were placed on top and any overlapping paper larger than the gel removed with a razor blade. Electrophoresis was performed at 80mA, 10 volts for 30-40 minutes for one blot or at 160mA for two.

After blotting, the nitro-cellulose was removed and washed briefly in phosphate buffered saline (PBS) and blocked in 30ml PBS, 2% w/v non-fat dried milk (Marvel) for 1 hour at room temperature. The membrane was twice washed with 30ml PBS for 3 minutes and immersed in 30ml PBS/Milk containing 0.1% v/v primary antibody and incubated at 4°C overnight. Excess antibody was removed by washing twice in 30 ml PBS, 0.05% Tween-20 for 3 minutes and twice in 30ml PBS for 3 minutes. 30ml PBS/Milk containing 0.1% v/v secondary antibody (conjugated to horse radish peroxidase HRP) was added to the membrane and incubated in for 40 minutes at room temperature. Finally the membrane was washed twice in 30ml PBS/Tween for 3 minutes followed by two 3 minute washes in 30ml PBS.

The antigen-antibody complex was located by immersing the membrane in 60ml HRP developing solution (0.3% w/v 4-chloro-1-naphthol, 8mM Tris-HCl pH 7.5, 125mM NaCl, 0.015% v/v H_2O_2) and fixed in water.

2.3 MICROBIOLOGICAL TECHNIQUES

2.3.1 Small scale preparation of plasmid DNA

Small quantities (<5mg) of pure plasmid DNA were obtained using a Miniprep SNAP kit (In-Vitrogen). This plasmid DNA was of a suitable standard for reliable DNA sequencing. The procedure recommended by the manufacturer was followed exactly.

2.3.2 Medium scale preparation of plasmid DNA

Medium sized preparations (10-25mg) of plasmid DNA were obtained using the a Maxiprep kit. The method recommended by the manufacturers (Promega) was followed exactly.

2.3.3 Large scale preparation of plasmid DNA

This procedure was performed to obtain between 25-100mg of plasmid DNA. 150ml of 2xYT containing the appropriate antibiotic and 2% w/v glucose was inoculated with a bacterial plaque containing the desired plasmid. The culture was grown at 37°C, 250 RPM overnight in a shaker incubator. The cells were harvested at 7,500 x g for 15 minutes and the pellet resuspended in 50ml STE (100mM NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA) and centrifuged again at 7,500 x g for 15 minutes. The pellet was resuspended in 9 ml of solution I (50mM glucose, 25mM Tris-HCl pH 8.0), 1ml of fresh lysozyme (10mg/ml) and the suspension gently mixed for 2 minutes. 20ml of solution II (0.2N NaOH, 1% w/v SDS) was added, the suspension gently mixed and incubated at room temperature for 10 minutes. 10ml of ice-cold solution III (4M potassium acetate pH 5.5, 11.5% v/v acetic acid) was added to the flask and the mixture inverted several times and incubated on ice for 10 minutes.

The lysed cells were centrifuged at 6,000 x g for 15 minutes (no brake to disturb the pellet). The supernatant was filtered through gauze into a clean flask. 0.6 volumes of isopropanol was added to the solution and left to incubate at room temperature

for 10 minutes. The solution was centrifuged at 6,500 x g for 15 minutes at room temperature, the supernatant discarded and the precipitated DNA allowed to air dry. The pellet was resuspended in 2ml TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). 2ml of 6M LiCl was added and the solution incubated on ice for 10 minutes followed by centrifugation at 8,000 x g at 4°C for 10 minutes. The supernatant was decanted to a sterile tube and the DNA ethanol precipitated (2.2.x). The DNA pellet was resuspended in 1ml TE buffer and 10µl RNase (10mg/ml in 10mM Tris-HCl pH 7.5, 15mM NaCl) was added and incubated at 37°C for one hour. The RNase was inactivated by incubation at 75°C for 10 minutes. The DNA was ethanol precipitated and resuspended in 0.5ml TE buffer.

2.3.4 Preparation of competent cells

50ml of bacterial culture was grown in LB to mid-log phase (A_{550} = 0.4) at 37°C in a shaking incubator (250 RPM). The culture was centrifuged at 3500 x g for 10 minutes and the supernatant discarded. The cell pellet was resuspended in 5ml ice-cold sterile 50mM $CaCl_2$, 10mM Tris-HCl pH 8.0 and placed on ice for 15 minutes to become competent. The cells were again centrifuged at 3500 x g for 5 minutes and the cell pellet resuspended in 15ml 50mM $CaCl_2$, 10mM Tris-HCl (pH 8.0), 15% (v/v) glycerol. The cells were then aliquoted in volumes of 250µl in pre-cooled eppendorfs and then snap frozen in an acetone/dry ice bath. The cells were stored at -70°C.

2.3.5 Transformation of DNA by Heat Shock

A 50µl aliquot of competent cells was mixed with approximately 25ng of the DNA to be transformed in a pre-chilled 1.5ml Eppendorf tube. The mixture was incubated on ice for 30-60 minutes after which it was heated in a water bath at 42°C for exactly 2 minutes then chilled on ice for 2 minutes. 750ml of LB media was added and the cells allowed to recover for 45 minutes at 37°C. After recovery the cells were centrifuged at 5,000 x g and the pellet resuspended in 50µl of LB media. The cells were plated onto LB agarose plates containing appropriate antibiotic and 2% w/v glucose and allowed to incubate at 37°C overnight.

2.3.6 In-well screening for recombinant plasmids

Recombinant plasmids were identified using an Epicentre Screening Kit based upon the protocol of Sekar, V. (1987). This is based upon differential migration of supercoiled plasmids that may or may not contain an inserted fragment. Therefore the inserts must be of a sufficient size to allow for visual discrimination (i.e. >100bp). A 1% agarose gel was prepared containing sufficient wells for the number of colonies screened. Once the gel was prepared, 15ml of Protoplast buffer was added to a number of microtitre wells in accordance to the number of colonies screened. Each colony, at least 2-3mm in diameter, was removed using a sterile toothpick and vigorously mixed into the protoplast buffer in the well. When completed the microtitre plate was incubated at room temperature for 10 minutes to allow protoplast formation. While the cells were incubating, 6ml of lysis buffer was pipetted into an appropriate number of wells of the agarose gel. The cell protoplast mixture was added to the lysis filled wells and allowed to incubate for 15 minutes. Before electrophoresing, molecular weight markers and a sample of supercoiled plasmid not containing any insert were loaded onto the gel. Sample containing the insert migrated more slowly than the control plasmid that contained no insert.

2.3.7 PCR screening for recombinant plasmids

If the size of the insert was insufficient or the plasmid used was not wild-type, PCR screening of cell plaques was used. A toothpick was used to pick *E. coli* plaques from the plate and were transferred into 20µl of sterile H_20. The cells were incubated at 94°C for 10 minutes to lyse the cells, after which 1µl was taken and added to the following reaction mixture: 2µl of x10 stock dNTP (300m), 2µl of 10x Thermopol buffer (100mM KCl, 200mM Tris-HCl pH 8.8, 100mM $(NH_4)_2SO_4$, 0.1% Triton X-100), 0.2µl of BSA (10mg/ml), 12.55µl H_20 and 0.25µl of Taq DNA polymerase (5 U/µl) were added to 0.2ml thin-walled tubes. The PCR cycle used was dependant of the oligonucleotide primers used for screening. 5µl of the reaction product was analysed on a 1% agarose gel.

2.3.8 Maintenance of bacterial strains and plasmids

Stocks of bacterial strains and *in situ* plasmids were prepared for long term storage at -70°C by mixing 800µl of a fresh overnight culture with 200µl sterile 80% v/v glycerol followed by snap freezing in dry ice. Short term storage strains were plated out onto LB agarose plates, (containing 2% w/v glucose + antibiotic), sealed with Nescofilm and stored in the dark at 4°C.

CHAPTER 3

3 MOLECULAR MODELLING OF 20G9 FV

3.1 INTRODUCTION

The aim of this thesis is to take a rational approach to altering the active site of the catalytic antibody 20G9 guided by a model of the combining site. Rational alterations are known as *protein engineering*. Protein engineering involves mutating the gene of an existing protein in an attempt to change its structure or function in a predictable way. To make alterations based simply upon the primary sequence of a given protein is hopelessly simplistic. Protein engineers have frequently been surprised at the range of effects upon a protein's structure and function caused by a simple point mutation when the aim was to change a specific function or property. Protein engineering must take into account the rules of protein stability and the energetics of catalysis. However, the speed of structural elucidation is slow when compared to the relatively rapid mutagenesis experiments. Solving the structure of an antibody generally takes up to three years including expression, biochemical characterisation, purification and crystallisation. Ultimately the protein may not even produce crystals satisfactory for X-ray studies. Thus structural elucidation is the rate determining factor in the turnover of the protein engineering cycle [Blundell and Sternberg, 1985]. An alternative method to crystallography is NMR, a considerably more rapid method of structure determination but is currently limited by an upper molecule size limit of approximately 30kDa where high-resolution structural information is required. This has forced the rapid evolution of molecular modelling methods capable of rapidly producing a structure but which is still very much prediction and therefore subject to program design errors.

3.1.1 Homology modelling

Predicting protein folding is difficult because of the enormous task of searching through all the possible conformations of a polypeptide chain to find those with the lowest energy. This can be reduced by taking account of sequence homology. X-ray crystallographic studies of protein families have demonstrated that proteins with homologous sequences also have similar folds. This provides the basis for homology modelling [Browne *et al*, 1969]. The first step in homology modelling was the alignment of known sequences for a family of proteins containing gaps in the sequence [Needleman and Wunsch, 1970]. Extensive alignment allowed structurally conserved regions (SCR's) and variable regions (VR's) to be assigned. It is thought that SCR's are usually located in the protein core and can be predicted with relative confidence whereas the VR's are generally surface located and are much harder to predict. Because VR regions have a poorly defined secondary structure they pose one of the main obstacles in predicting the folding of variable fragments of antibodies.

3.1.2 Modelling antibody combining sites

Padlan and Davies were the first to attempt modelling of antibody combining sites [Padlan *et al.*, 1976] when relatively few antibody structures were known. Recognising that the SCR's that lay within the β-sheet framework regions of antibody variable domains were relatively easy to model by homology, Padlan reasoned that the CDRs could possibly be modelled in the same way based upon other antibody CDR structures. This was further developed by Rees when more antibody structures became available [Darsley and Rees, 1985; de la Paz *et al*, 1986]. In 1986 Chothia & Lesk introduced the knowledge based concept of "key" residues [Chothia *et al.*, 1986] using a crystallographic database of antibodies. These 'key' conserved residues are defined by their packing, hydrogen bonding and ability to assume unusual Ψ, ϕ and ω torsional conformations and are thought to be responsible for the main conformations of the hypervariable regions. This allowed a classification of CDRs of similar lengths into a series of "canonical" structures. The canonical structure approach has succeeded in modelling five of the six CDRs (L1, L2, L3, H1, H2). However CDR H3 doesn't fit into a canonical class due to

the loops greater variability in sequence, length and structure in comparison to the other CDRs. This is due to the fact that the H3 CDR gene sequences are situated at the position where the J and D segments join. In addition, the residues coded at the beginning of the J_H genes may not even be present in the final structure. This variability is compounded by somatic diversity resulting in a loop that varies greatly in both composition and size when compared to the other five CDRs making prediction much more difficult but not impossible [Chothia and Lesk, 1987].

3.1.3 Antibody topology and composition

There are three shape classes which describe the overall topography of the antibody combining site based upon antibody x-ray crystallographic structures: i) cavity antibodies ii) groove antibodies iii) planar antibodies [Wang *et al.*, 1991; Webster *et al*, 1994]. The topography type is thought to be connected to the type of antigen. Smaller antigens have been associated with cavity antibodies [Rini *et al.*, 1992] and are buried in a hydrophobic pocket in the combining site. Larger antigens appear to bind over a much greater surface [Sheriff *et al.*, 1987] forming many interactions and are generally associated with planar or groove antibodies. Peptides or larger haptens generally result in groove antibodies.

Differences between the sidechain compositions of framework and CDRs are commonly observed. Studies have shown [Mian *et al.*, 1991; Padlan, 1990] that CDR sidechains are generally more exposed than of those in the framework and that there are more bifunctional residues such as His, Tyr or Asn in the CDRs. This contributes hydrophobic character to the loops as well as increasing their ability to engage in hydrogen bonding. Furthermore there is a high frequency of exposed hydrophobic residues. It is thought that these hydrophobic sidechains are involved in antigen binding whereas the charged sidechains mainly account for antigen specificity [Colman, 1988].

3.2 ANTIBODY MODELLER, A COMBINED ALGORITHM

The immunoglobulin domain modelling program, *AbM*, uses a **C**ombined **A**ntibody **M**odelling **Al**gorithm or CAMAL. The undertaking of producing an antibody model is approached by reducing the problem into a series of smaller tasks [Martin *et al.*, 1989, Pedersen *et al*, 1992]. Each task can then be systematically tackled applying the appropriate method from a panel of options.

3.2.1 Building the framework

Antibody framework regions consist of conserved sequences that form β-barrel structures. The antibody model begins with the building of a framework starting with sequence entry by the user. This sequence is aligned against a databank of heavy and light chain antibody sequences of known structure. Alignment is achieved by inserting gaps until the sequences are all of same length. As framework regions of antibodies are highly conserved, alignment is usually easy. The program compares light and heavy chains against the sequences within the database. A similarity calculation is performed to find the database sequence that most closely matches the conserved framework regions of the target sequence. Gaps in the aligned sequence that are paired with conserved framework residues are very heavily penalised by the calculation. Finally a candidate framework for the model is proposed based upon the scores.

The antibody structure database used by *AbM* contains files of the co-ordinates and pre-aligned sequences of the light and heavy chains of known antibodies. The contents are a selection of full and pre-release Brookhaven Protein Databank structures solved to a high resolution. Sequences in the database have been pre-aligned to help maximise sequence identity and to reflect regions of most structural variation. In addition, it contains average co-ordinates from known antibody structures and are representative of the most highly conserved residues of the β-barrel. This "core" β-barrel is used as a template onto which the selected sequence is fitted.

3.2.2 Construction of CDR loops

CDR loops of the target sequence are examined against canonical sequence patterns of a database. This allows assignment of a canonical class for each loop. Once classified the target loop is assessed for similarity towards antibody database loops belonging to the same class. Those loops with greatest homology are then used to the model the target loop. This is only applied to the three light chain loops, (L1, L2, L3) and the two heavy chain loops, (H1 and H2) as no canonical classes currently exist for the H3 loop.

Any loops of CDRs L1-3 and H1-2 that do not belong to a canonical class are constructed using CAMAL. The build method of non-canonical loops is decided based upon the number of residues within it. A loop of up to three residues is built by conformational searching whereas loops of between four and seven residues are built by Cα database searching. Loops with greater than seven residues are built using a combination of conformational and Cα database searching.

Conformational and Cα database searches. Conformational searching is performed on loops with three or less residues, as computational time increases exponentially with the length of the loop. Using the conformational search algorithm, CONGEN [Bruccoleri and Karplus, 1987], a series of candidate backbone conformations are produced. CONGEN achieves this by saturation of available conformational space by rotation about the ϕ,ψ and ω torsion angles under restrictions based upon a form of the Ramachandran plot [Ramachandran *et al.*, 1963]. This plot provides a range of torsion angles to which amino acids are normally restricted by a specific energy threshold. This allows elimination of energetically disfavoured structures and backbone torsion's that do not exist naturally. Side chains are generated around all possible χ angles and those sidechains with the lowest energy level recorded. This is repeated for each side chain beginning with the N-terminus residue until a constant energy level is

reached. The energy levels of each completed loop are filtered using *Eureka* with loops of lowest energy subsequently added to the working model. *Eureka* is a force field used for energy determination [Dauber-Osgusthorpe *et al.*, 1988] of peptides and proteins.

Generation of loops of between four and seven residues in length is achieved by searching the Cα database. This database contains Cα-Cα distances and backbone angles derived from protein structures in the Brookhaven Protein Databank. It only contains proteins with greater than 20 bonded amino-acids containing well defined Cα-Cα distances at resolutions of 3Å or better. The target loops are selected upon Cα distances and have side chains added using CONGEN. Each candidate loop is added to the working model in turn. Five working models are created with each containing one of the five loops found to have the lowest energy. A Structurally Determining Residue (SDR) Filtering method [Sutcliffe, 1988] is applied to each of the working models to assess their resemblance to the database loops and preference of loop geometry found within nature.

Loops with greater than seven residues are generated using a combination of conformational and Cα database searching. The basic loop backgrounds are usually found by searching the Cα database but with reduced hits due to the large loop size. The basic loop backbone is positioned onto the model and the top 5 residues of the loop deleted and reconstructed using CONGEN. This allows saturation of the conformational space available to each segment by varying the Ψ and ϕ torsion angles on a grid. The resulting loop structure has its base residues derived from the protein structure database and those at the top based upon saturation of conformational space. Side chains are added using CONGEN and the loop screened using SDR filtering.

3.2.3 Assessment and minimisation of the model

Once a basic model has been produced there follows a process of objective evaluation and minimisation. No minimisation is applied during the building of

the basic model as this requires much computational time for calculations with relatively little observed benefit. As the framework is directly derived from crystal co-ordinates it is relatively accurate and so it's more effective to assess and minimise the whole model after construction. However the loop backgrounds, especially those constructed by non-canonical methods and the side chains are less accurate and so minimisation is an important in refinement of the model.

3.3 PRODUCING THE 20G9 MODEL

3.3.1 Sequence entry

The sequence of the light and heavy variable regions were aligned into Kabat [Kabat *et al.*, 1991] format before entry into *AbM* (Figure 4.3). This simplified the manual alignment step to the displayed Fv sequences within *AbM*.

On execution of *AbM*, the algorithms described above process the sequences. The model of the Fv is produced in the form of a PDB (Protein Data Bank) file, which contains structural information of the model. It contains the sequence number of each atom, the atom type, the type of residue of which it is part, the sequence number of the residue and the x, y, z co-ordinates of the atom within the whole structure (Figure 3.1). These files can be displayed using software of various complexities, enabling the display of different image forms based upon the same data, i.e. RasMol (Sayle, R. GlaxoWellcome R&D, Greenford, UK), Insight (Biosym Technologies, San Diego, USA), MolViewer [Hartshorn, 1996] or MolScript [Kraulis, 1991]. The benefits of this are that various software packages have features useful for different tasks. Some enable the user to rapidly examine the structure without too great a drain on the computer processor, while others allow production of images suitable for publication but far too complex to allow rapid 'hands-on' on-screen rotation. The details of atoms and residues contained within the file are important and can be easily identified within the on-screen image by simply pointing and clicking at the point of interest. This allows rapid identification of specific residues that are potentially situated within important positions within a structure. The numbering scheme of the residues within the PDB

file exists as a single string and are identified with a different numbering scheme to that of the Kabat alignment (Figure 3.2).

SERIAL:	the atom serial number
STRING:	the atom name and type
RESIDUE:	the residue name for the entry (4 characters)
SEQUENCE:	the entry sequence number
X Y Z:	the atomic co-ordinates

```
ATOM   133  C     TYR  28    -1.630     -1.423      0.847
ATOM   134  O     TYR  28    -2.941      2.214     -0.925
ATOM   135  N     TYR  28    -0.288     -0.133      0.321
```

Figure 3.1 Each atom within the PDB file is numbered, labelled by atom type, the residue name to which the atom belongs, the residue number within the protein sequence and 3 dimensional atomic co-ordinates. These data allow the file to be displayed using different pieces of software which utilise various aspects of the data contained within the file all based upon the details and co-ordinates given.

```
                                  [L1]
DIVMTQAAPSVIVTPGESVSISC RSSKSLLHSYGNTYLY WFLQRPGQSPQVLIY
|        |         |          |          |         |
1       10        20         30         40        50

  [L2]                                        [L3]
RMSNLAS GVPDRFSGSGSGTAFTLRISRVEADDVGVYYC MQYLEYPYT FGGGTKLELK
     |          |         |        |          |          |
    60         70        80       90        100        110

                                  [H1]
EVQLHEPGAALVRPGASVMLPCRASGYSFT SFWMN WVRQRPGQGLEWIG
     |         |         |           |         |
    120       130       140         150       160

       [H2]
RIHPSDNRTRLNQKFKD KATLTVDKSSSTAYMQLISPTSEDSEVYYCAR
      |         |          |         |         |
     170       180        190       200       210


[H3]
EGGAY WGQGTLVTVS
       |
      220
```

Figure 3.2 Continuous residue numbering of the 20G9Fv PDB file.

AbM successfully produced a model of 20G9 Fv and was able to base the L1, L2, L3 and H1, H2 loops upon known canonical structures within the search database. The H3 cannot be based upon a canonical class and so candidate loops were constructed by the use of the Cα database with sidechains added using CONGEN (Table 3.1).

Table 3.1 Canonical classification of 20G9 CDRs. Canonical loops existed for five of the six loops. However no canonical determinants are known for CDR H3 thus this loop was modelled using the Cα database and sidechains added using CONGEN.

CDR	Sequence	CDR Length	Canonical class	Modelling method
L1	RSSKSLLHSYGNTYLY	16	4	canonical
L2	RMSNLAS	7	1	canonical
L3	MQYLEYPYT	9	1	canonical
H1	SFWMN	5	1	canonical
H2	RIHPSDNRTR	10	3	canonical
H3	EGGAY	5	NA	CONGEN

3.3.2 Examination of the 20G9 Fv model

The 20G9 Fv antibody binding site appears as a long groove containing a central cavity into which the hapten might bind (Figures 3.3 and 3.4). Small haptens conjugated to a large BSA molecule are often associated with antibodies that have a groove for their active site. Examination of the primary sequence and the model of 20G9 show that the light variable chain has a high number of tyrosine residues while the heavy variable chain possesses relatively few.

Figure 3.3 Molscript generated space filling model of 20G9 Fv generated by AbM. This side view displays a clear groove running through the combining site. Hydrophilic and hydrophobic patches are shown on the surface as red and blue respectively.

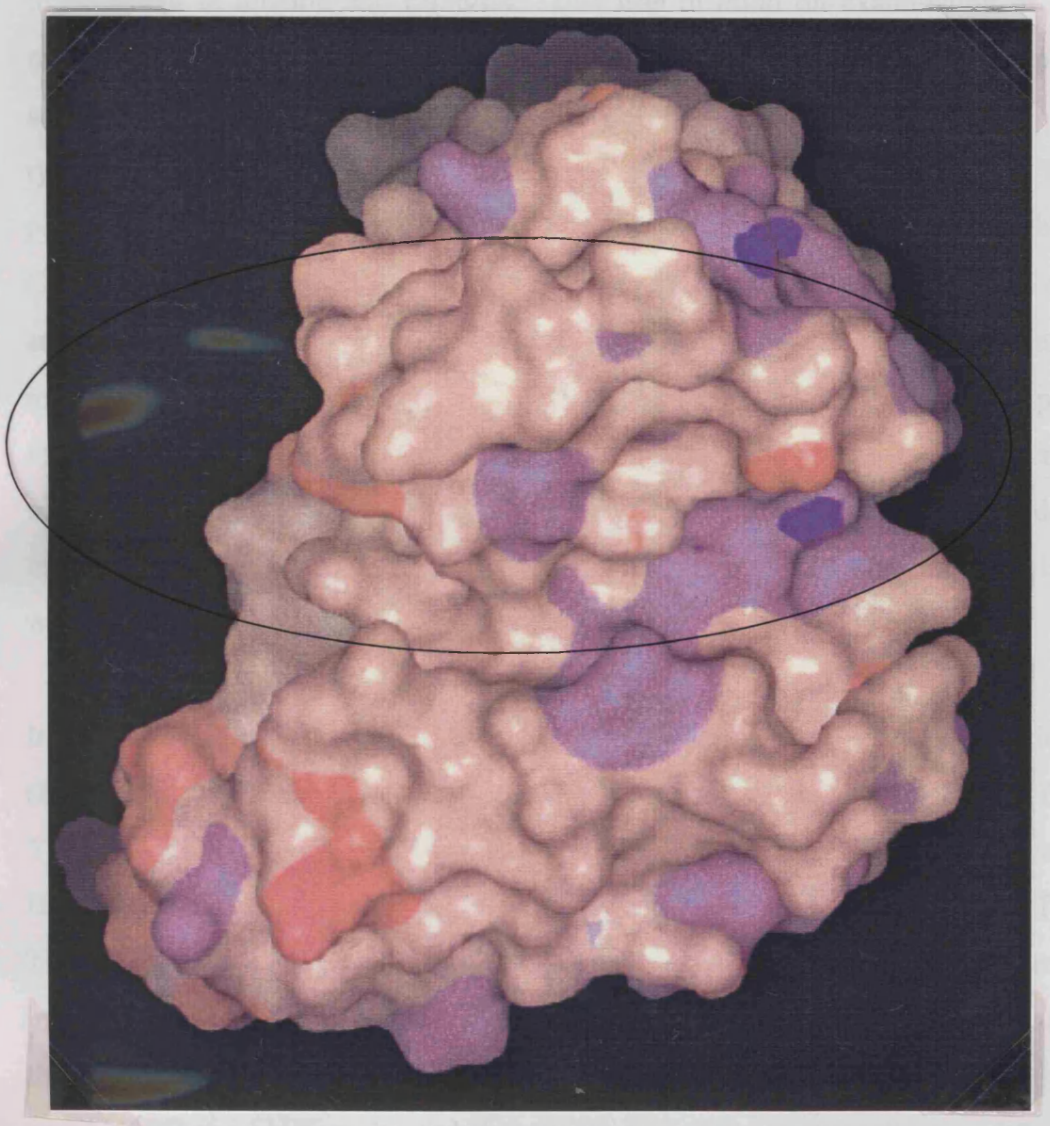


Figure 3.4 Molscript generated space filling model of 20G9 Fv generated by AbM. The view down into the binding again clearly shows the groove running through the combining site. Hydrophilic and hydrophobic patches are shown on the surface as red and blue respectively.

3.3.3 Targeting residues for mutation

Prediction of residues critical for binding and catalysis is a difficult task and must be based on structural evidence from the model, information from the kinetic characterisation of 20G9 and also on those residues commonly found in other catalytic antibodies. Previous kinetic studies on 20G9 implicated tyrosine involvement in binding and catalysis. This was in mind on examination of the model. At the time of the modelling, the latest mutagenic and crystallographic studies demonstrated critical residues in the active sites were frequently found to be tyrosine and arginine, again taken into account when selecting residues. The purpose of the model was to identify residues within the light and heavy chain CDR's or fringes (Figures 3.5 and 3.6) with particular attention paid to tyrosine and arginine residues proximal to the putative active site. Rotation of the wireframe model of 20G9 was used to gauge the relative positions of the residues with respect to the groove. Non-tyrosine/arginine residues with a close proximity to the active site were also selected. The residues were then graded primary or secondary mutations depending on the likelihood of involvement, in order to determine into which wave of mutations they would fall (Figures 3.7 and 3.8).

In total, 12 residues were identified of which six were graded primary mutations (PDB/Kabat numbering): R164/^{H}50Arg, R171/^{H}56Arg, R173/^{H}58Arg, Y96/^{L}91Tyr, Y99/^{L}94Tyr, Y101/^{L}96Tyr. Residues selected for secondary mutations were: H166/^{H}52aHis, D169/^{H}54Asp, Y33/^{L}28Tyr, Y37/^{L}32Tyr, Y39/^{L}34Tyr, Y54/^{L}49Tyr. The tyrosine and the arginine residues were selected because of their common involvement in abyzme catalysis, although they were not immediately proximal to the binding site. The histidine and aspartate residues were selected because of their proximity to the binding site and their generally reactive nature (Figures 3.8 and 3.9). A degree of flexibility was used when examining the residues on the model, selecting residues that were within a reasonable distance of the groove centre (approximately 15 **Å**), even when their orientation may not have been ideal.

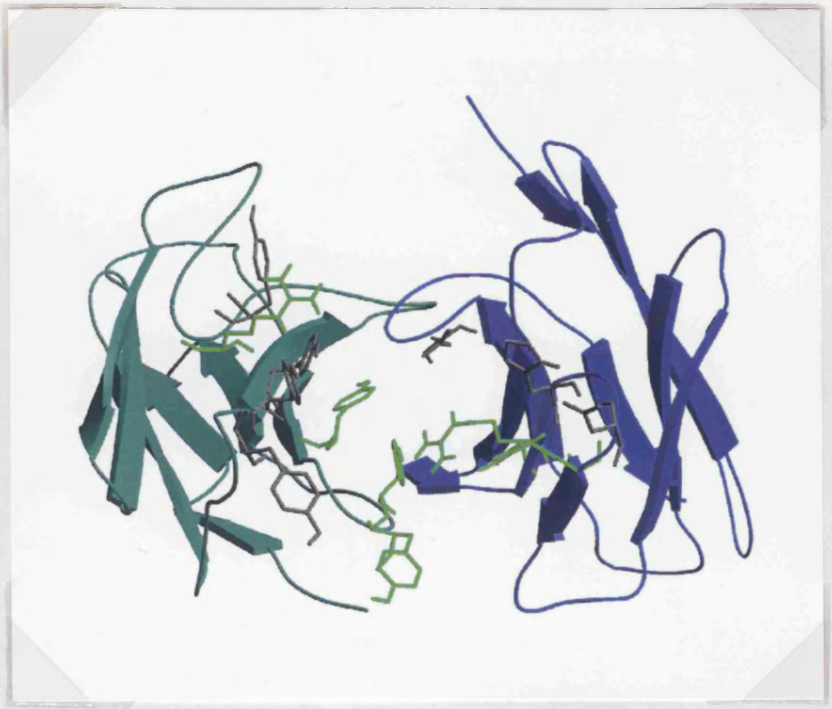


Figure 3.5 Molviewer ribbon representation of 20G9 Fv model with residues selected for mutagenesis studies displayed as wireframe. The residues highlighted in green are those thought most likely to be involved within the active site of the molecule and were selected for the primary round of mutagenesis studies. The remaining residues displayed in grey were selected for the secondary round of site-directed mutagenesis and are thought less likely to play an active role within the catalytic site of 20G9.

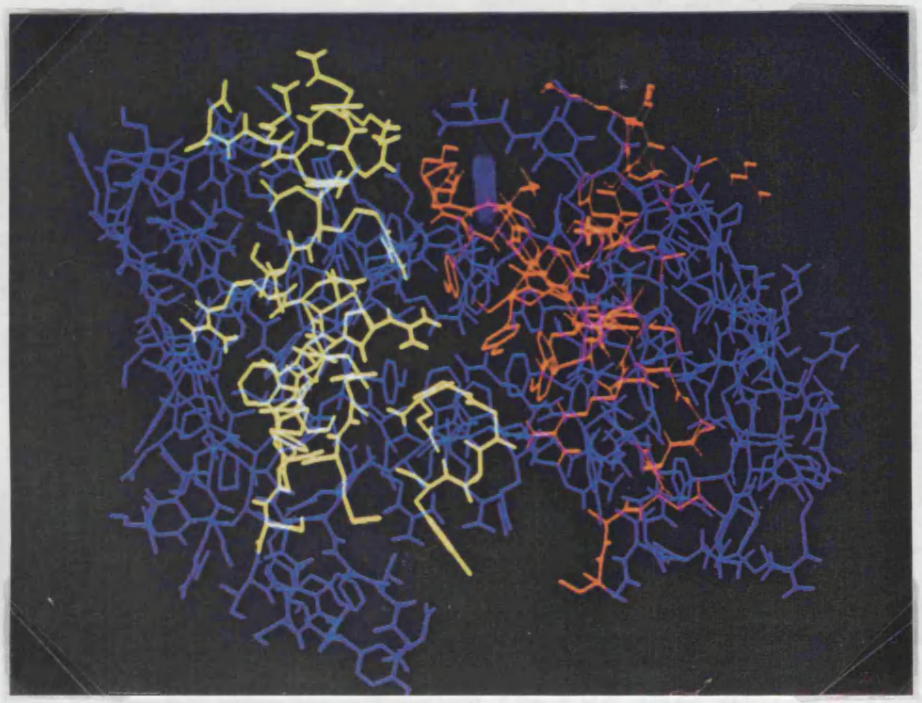


Figure 3.6 An Insight II wire frame representation of the model of 20G9 Fv which was generated using *AbM*. The view is directly down into the combining site of the antibody fragment. Light chain CDRs are in red and heavy chain CDRs in yellow.

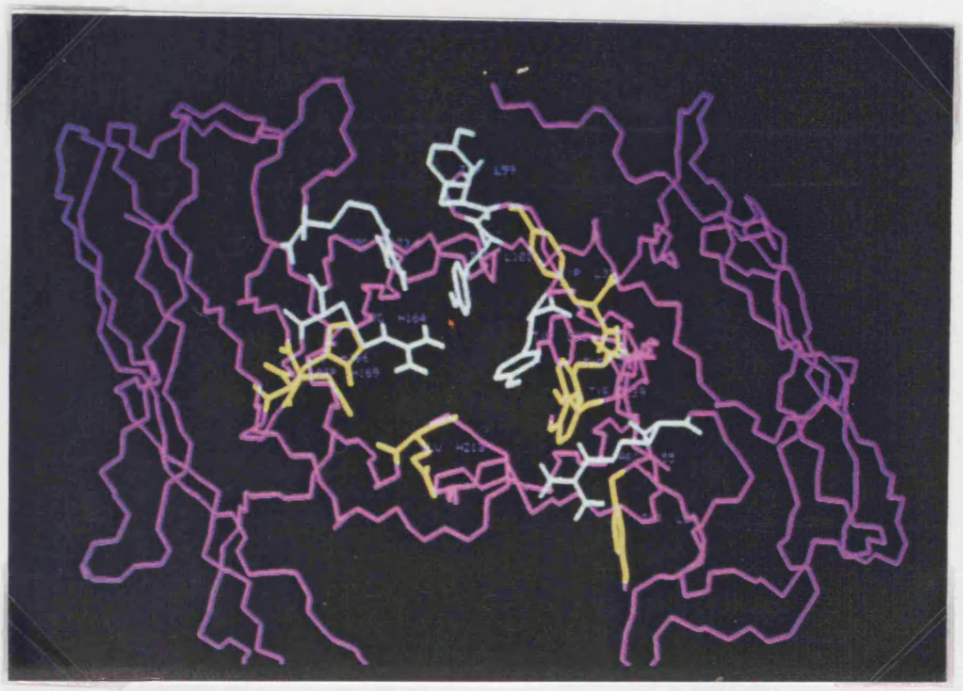


Figure 3.7 This wire frame representation of 20G9 Fv was used to examine residues thought likely to have a role in binding. Residues selected for site-directed mutagenesis are indicated: primary mutations (highlighted white), secondary mutations (highlighted yellow).

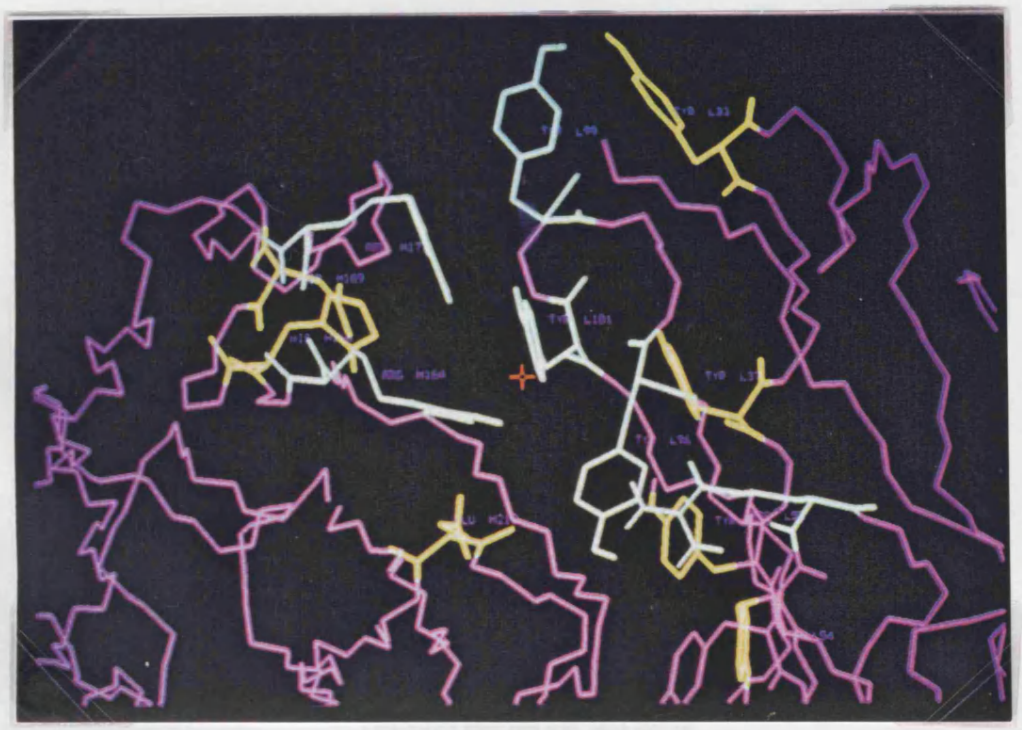


Figure 3.8 Insight II wire frame representation of 20G9 Fv. The closer view allows identification of the white primary residues (L99, L96, L101, H164, H171, H173) and the yellow secondary residues (L33, L37, L39, L54, H166, H169, H213). *[PDB numbering notation with a L or H denoting heavy or light chain]. *The residues H213 and L55 were marked in error and should not have been highlighted.*

3.4 DISCUSSION OF THE 20G9 MODEL

3.4.1 Integrity of the 20G9 Fv model

It is likely that the 20G9 Fv model is a good representation of the actual structure as its sequence presented no particularly difficult problems. Five of the six CDRs excluding H3 were selectable as canonical structures and as a result, their conformations should be good. The H3 loop is always the most difficult loop to model since it lacks a good canonical classification. The H3 loop of 20G9 consists of only 5 residues, making the task of modelling the loop relatively straightforward. Short H3 loops are generally modelled with accuracy's of around 1.5Å RMSD, sufficient to guide the present study [Rees *et al*, 1996].

3.4.2 Selection of residues

The final step was to identify residues that were part of or proximal to the active site of 20G9. These would then be mutated one by one to alanine, in an attempt to modify or eliminate activity. This will allow us to say with a degree of confidence that the selected residue is critical for catalysis. Initially, consideration was given to mutating the tyrosine residues to phenylalanine, to preserve the ring structure and thus minimise structural disruption. However, if the aromatic ring present in a given tyrosine was playing a role in ground state binding, or in product inhibition due to aromatic stacking, then this may not have been eliminated. The more radical mutation of tyrosine to alanine was selected to avoid any aromatic stacking effects. These experiments are described in the next chapter.

CHAPTER 4

4 SITE-DIRECTED MUTAGENESIS OF 20G9

4.1 INTRODUCTION

To understand the functions of genes and their cognate proteins it is valuable to have the means to alter them in some manner so that the mutated protein can be compared against the original protein. Before the existence of current molecular biology techniques, mutants were generated by treatment of organisms (e.g. bacteria) with mutagenic agents or with physical stress in order to modify the DNA. However, such crude methods had the disadvantage that even when phenotypic changes were observed, one could not be sure if this was due to changes in the gene of interest or to some other genetic or epigenetic event. In addition, even if the targeted gene was altered, the nature of the change (single/multiple base changes, deletion/insertion of DNA, etc) would often remain unknown. Today the isolation and study of single genes has become routine through continual progress in mutagenic techniques. Previously, mutagenesis was a major task requiring many DNA manipulations to create or eliminate restriction sites in order to alter a gene. It is now relatively simple to alter a specific base in a DNA sequence and consequently mutagenesis has become an important basic tool in genetic manipulation.

4.1.1 Single oligonucleotide mutagenesis

Site-directed mutagenesis can be carried out using the single oligonucleotide method [Gillam *et al.*, 1980; Zoller & Smith, 1983]. This method uses a single chemically synthesised oligonucleotide carrying a base mismatch towards the complementary sequence. The DNA sequence that is to be mutated must exist in a single stranded form which can be achieved using M13-based vectors or by denaturing double stranded into single stranded molecules. The oligonucleotide primes DNA synthesis and becomes incorporated into the resulting heteroduplex

molecule with one strand for the wild-type sequence and the other coding for the mutant. On transformation, the heteroduplex molecule separates into a homoduplex molecule and the resulting plaques contain a mix of wild type and mutant plasmids. Often the ratio of mutants to wild type is low due to the methyl-directed mismatch repair system of *E. coli.* Bacteria attempt to repair the non-methylated DNA of the mutant that is paired with the complementary sequence of the wild-type thus preventing mutations [Kramer *et al.* 1984a). In order to prevent this, the host strain must carry the gene mutations, *mut*L, *mut*S or *mut*H within its genotype, which prevent methyl-directed repair of mismatches.

To enhance selection of plaques containing the mutant plasmid, many commercial methods have been developed using antibiotic selection to select mutants. The single primer method described above is employed by the Altered Sites II site-directed mutagenesis system (Promega). The system uses a vector, pALTER-1 that carries an inactivated ampicillin resistance gene. The gene of interest is cloned into the vector and the plasmid double stranded DNA on exposure to alkaline conditions separates into single stranded DNA. The mutagenenic oligonucleotide and ampicillin resistance repair oligonucleotide primers are annealed to one of the strands, the 'template'. The primed template forms the mutant strand by the addition of DNA polymerase/DNA ligase. The double stranded construct must be transformed into modified *E. coli* bacterial strain (e.g. *mut*S) to prevent repair. Ampicillin selection allows the survival of only those bacteria transformed with the mutant plasmid carrying the repaired ampicillin gene. The mutated sequence is then restricted from the pALTER-1 vector and ligated into an appropriate expression vector. When using this method it is critical that the oligonucleotide carrying the mutant mismatch anneals and primes well against the template. The ampicillin gene repair oligonucleotide provided with the system normally anneals and primes efficiently regardless of the efficacy of other oligonucleotides. Subsequently, a double stranded construct with antibiotic resistance will be positively selected but may carry the wild type gene if the mutant oligonucleotide fails to bind to the template.

4.1.2 PCR Mutagenesis

PCR is commonly used for analysis of DNA sequences but is equally useful for *in vitro* synthesis of DNA. The technique when used in a controlled manner can produce specific alterations within amplified DNA by the intelligent use of oligonucleotide primers. The 5'-end of the oligonucleotide can be designed to encode any sequence provided the 3'-end of the oligonucleotide contains sufficient complementary DNA to effectively bind the template. During the PCR amplification process, oligonucleotides can then be incorporated into the product DNA. This allows the production of chimeric DNA that carries the sequence of the template DNA centrally but with the designed DNA at the end(s). This allows site-directed mutagenesis to be rapidly and easily performed on the target DNA, [Scharf *et al.* 1986].

If mutations are desired within the centre of the DNA product, this can be achieved by the design and production of long primer oligonucleotides reaching into the middle of the fragment to be mutated. This has the drawback of high cost of lengthy oligonculeotides and the greater chance of the introduction of wrong nucleotides into the oligonucleotide during its production, (the longer the priming oligonucleotide, the greater the chance of PCR errors occuring). This can be overcome by using the overlap extension method. This allows mutations to be introduced in the centre of DNA fragments by designing oligonucleotides so that PCR amplified products overlap with other PCR amplified products. A subsequent reaction can be performed with the overlap priming for extension from which DNA polymerase creates the entire recombinant molecule (Figure 4.1), [Higuchi *et al.*, 1988; Ho *et al.*, 1989). This method superseded the DNA ligase method of fusing together two DNA fragments and can be applied to produce specific changes, insertions and deletions within fragments of DNA. Mutagenesis by overlap extension is achieved by introducing bases changes into the oligonucleotides where the two PCR products overlap. It is entirely possible to introduce the base changes into one oligonucleotide only but this may cause a potentially difficult extension reaction to become impossible. Only by designing two oligonucleotides that are entirely complementary on the overlap and of sufficient length will the extension reaction become optimal (Figure 4.2). Using PCR overlap extension for

mutagenesis represents a significant improvement over the single-oligonucleotide vector systems described above due mainly to its efficiency and simplicity. Relatively few steps are required to produce mutant inserts that can be sequenced to eliminate undesired mutations. Once satisfied that there are no errors (by sequencing the PCR product) , the insert may be ligated into the vector, eliminating the necessity of screening for mutant constructs.

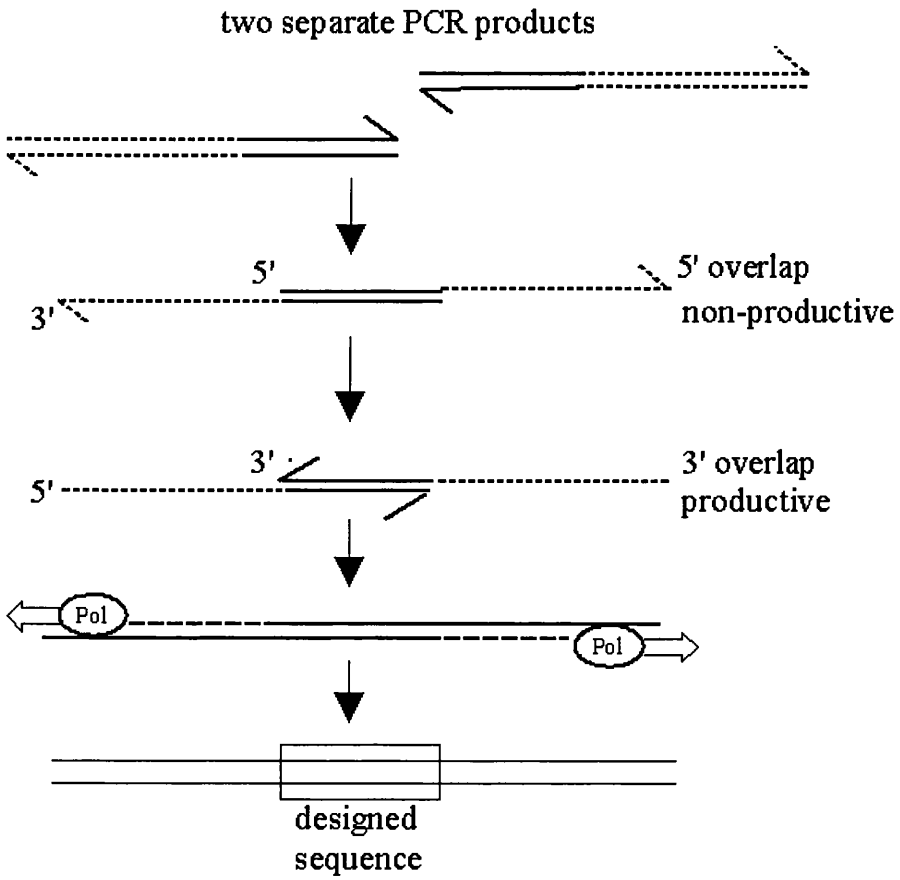


Figure 4.1 Diagram of PCR overlap extension. Two separate DNA fragments can be combined if a sequence overlap exists between the two fragments. These overlaps will prime for extension of DNA when DNA polymerase is added resulting in the production of a single recombinant molecule. This technique can be used to introduce changes anywhere within a DNA sequence through the careful design of oligonucleotides.

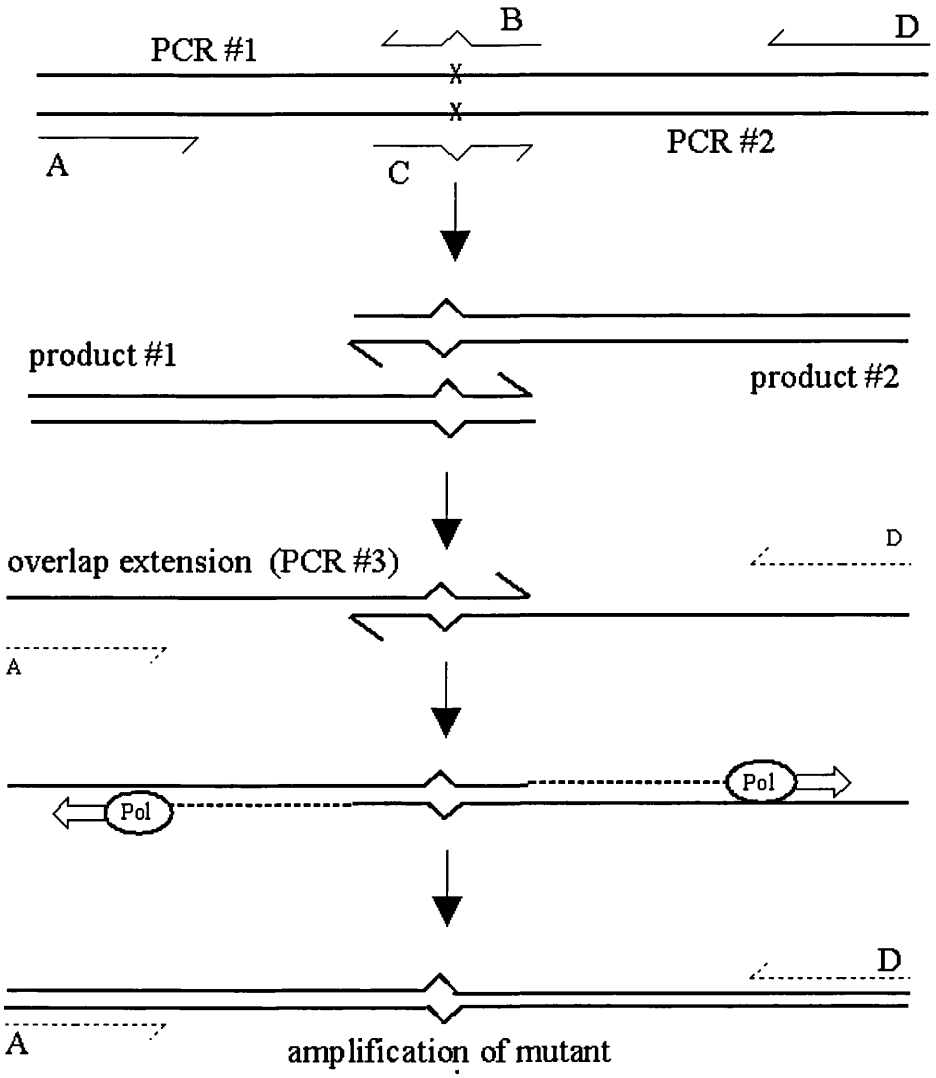


Figure 4.2 Site-directed mutagenesis by overlap extension. Two separate reactions are performed. PCR#1 contains oligonucleotide A and B and PCR#2 contains oligonucleotide C and D with each reaction incorporating a mutagenic oligonucleotide. Once completed, the PCR products are cleaned and placed into a third reaction containing the two end oligonucleotides (A & D) that were used in the first round of reactions. Overlap extension of the products occurs and the recombinant molecule is amplified using oligonucleotides A & D. These two oligonucleotides may be used to introduce restriction sites into the recombinant molecule.

It is critical to sequence all PCR generated products as errors can be introduced during the PCR process. The two main factors that introduce errors are:

i) the number of PCR cycles
ii) the polymerase used.

Misincorporations early in the amplification cycle result in propagation of that error through subsequent cycles. This DNA would represent a significant portion of the final product and would probably be cloned into the vector. This problem cannot be eradicated but can be minimised. By adding more target template to the reaction, the number of cycles necessary to produce the required quantity of product is reduced thus reducing the opportunity for error.

The different DNA polymerases used in PCR vary in their capacity to introduce errors into products (Table. 4.1). The fidelity of a polymerase is directly proportional to its cost, with those more prone to errors found at the lower end of the market. For site-directed mutagenesis the highest fidelity polymerases available are Vent DNA polymerase (NEB) and Expand polymerase (Boehringer-Mannheim, Mannheim, Germany). Their low error rates are due to a 3'-5' proof-reading exonuclease property. Each base is checked against the template, which ensures that no misincorporations are retained, but excised as detected. This results in an error rate of 5.7×10^{-5} compared to that of the commonly use Taq DNA polymerase which has a rate of 2.8×10^{-4}. A common drawback with Vent polymerase is that PCR is highly sensitive to reaction conditions and many reactions fail to produce product or produce poor product due to non-specific binding resulting in heavy smearing. Careful design of oligonucleotides and optimisation of reaction conditions are necessary if a product of acceptable standard is to be produced.

	Vent DNA Polymerase	Deep Vent Polymerase	Expand DNA Polymerase	T4 DNA Polymerase	Taq DNA Polymerase
Error Rate($x10^{-6}$)	57	65	8.5	< 1	285
5'-3' Polymerase	+	+	+	+	+
3'-5' Exonuclease	+	+	+	+	-
Thermal stability	+	++	+	-	+

Table 4.1 The fidelity of polymerase varies widely and is often dependent on the presence of proof-reading activity enabling the discovery, excision and replacement of misincorporated nucleotides. Relatively few polymerases are suitable for use in PCR with those used requiring heat stability of temperatures up to 100°C. (Figures obtained from Boehringer Mannheim and NEB data sheets).

4.2 DESIGN OF OLIGONUCLEOTIDES FOR MUTAGENESIS

Mutation positions were selected using the 20G9 Fv model as a guide (Figure 4.3 & section 3.4.2). It was necessary to design oligonucleotides for use with PCR site-directed mutagenesis. This required the design of two complementary primers for each mutation in order to produce the upstream and downstream overlapping fragments for overlap extension.

V_L (κ)

```
1        10         20    24 27abcde28    34 35  40        49
|         |          |     |  |||||||      | |    |         |
DIVMTQAAPSVIVTPGESVSISC RSSKSLLHSYGNTYLY WFLQRPGQSPQVLIY

50    56
|     |
RMSNLAS

57 60        70         80       88 89      9798100    107
|  |         |          |        | |        | | |       |
GVPDRFSGSGSGTAFTLRISRVEADDVGVYYC MQYLEYPYT FGGGTKLELK
```

V_H

```
1        10         20         3031  3536  40        49
|         |          |          | |   | |   |         |
EVQLHEPGAALVRPGASVMLPCRASGYSFT SFWMN WVRQRPGQGLEWIG

5051 52a b  53     60    65 66 70         80
|  |  |  |  |       |     | |  |          |
R  I  H  P  SDNRTRLNQKFKD KATLTVDKSSSTAYM

82abc83     90  94 95 96  100 101 102 103      112
 |||||       |   | |  |   |   |   |  |         |
QLISPTSEDSEVYYCAR E   G   G   A   Y  WGQGTLVTVS
```

Figure 4.3 Mutation of residues based upon the 20G9 *AbM* model. 1° and 2° residues targeted for mutation are underlined. Residues are numbered using the Kabat [Kabat *et al.*, 1991) numbering scheme.

The sequences of the V_L and V_H regions of 20G9 were entered into DNA Stryder v1.0 (C. Merck, Dept. of Biology, Institute of Fundamental Research, Commission of Atomic Energy, France) and analysed for amino acid codon usage (Appendix).

GCT was the dominant codon used for alanine residues within 20G9. Therefore the alanine mutants that were introduced into 20G9 used the codon GCT rather than GCC, GCA or GCG to maintain codon usage consistency within the gene. Oligonucleotides of approximately 21 nucleotides were designed to anneal at least nine bases on either side of the mutated codon. Each oligonucleotide was checked by using DNA Amplify (Bill Engels, University of Wisconsin, Genetics, Madison, WI 53706, USA) in order to eliminate any non-specific binding and avoid dimerisation of the oligonucleotides to themselves or any partner oligonucleotides. DNA Amplify also allowed calculation of the PCR fragment size and its probable intensity, which was useful in verifying the success of each PCR.

The primary mutations were designed and synthesised first. Each oligonucleotide was 21 nucleotides long with the codon encoding the mutated residue flanked on both sides by 9 bases. Each oligonucleotide primer was checked using DNA Amplify and checked against the pKKpelB/20G9WT construct to minimise any non-specific binding before their synthesis. It was found that the oligonucleotide, MR58A3, dimerised to itself. In order to prevent this, three bases were deleted from the 3' end, and the 5' end was increased by the addition of 5 bases. Oligonucleotides were custom synthesised and HPLC purified by Perkin-Elmer-Applied Biosystems, Warrington, UK., and were supplied lyophilised. The oligonucleotides were reconstituted in sterile MilliQ water and quantified spectrophotometrically at A_{260} and diluted to 50pM for use with Vent DNA polymerase or to 10pM if used with *Taq* DNA polymerase. The diluted oligonucleotides were aliquoted in suitable quantities and stored at -20°C. In total, 12 oligonucleotides (Table 4.2) were produced for the primary wave of mutations. The primers necessary for the secondary wave of mutations were designed (Table 4.3) but not synthesised pending the characterisation of the first series.

$Y^{L91}A$	Y=TAT	A=GCT	
MY91A5:	5' TA TTC TAG **AGC** TTG CAT ACA 3'	20mer	$T_m = 54°C$
MY91A3:	5' TGT ATG CAA **GCT** CTA GAA TA 3'	20mer	$T_m = 54°C$
$Y^{L94}A$	Y=TAT	A=GCT	
MY94A5	5' CGT GTA CGG **AGC** TTC TAG ATA 3'	21mer	$T_m = 62°C$
MY94A3	5' TAT CTA GAA **GCT** CCG TAC ACG 3'	21mer	$T_m = 62°C$
$Y^{L96}A$	Y=TAC	A=GCC	
MY96A5	5' TCC GAA CGT **GGC** CGG ATA TTC 3'	21mer	$T_m = 66°C$
MY96A3	5' GAA TAT CCG **GCC** ACG TTC GGA 3'	21mer	$T_m = 66°C$
$R^{H50}A$	R=AGG	A=GCT	
MR50A5	5' AGG ATG AAT **AGC** GCC AAT CCA 3'	21mer	$T_m = 52°C$
MR50A3	5' TGG ATT GGC **GCT** ATT CAT CCT 3'	21mer	$T_m = 52°C$
$R^{H56}A$	R=AGA	A=GCT	
MR56A5	5' TAA CCT AGT **AGC** ATT ATC GGA 3'	21mer	$T_m = 48°C$
MR56A3	5' TCC GAT AAT **GCT** ACT AGG TTA 3'	21mer	$T_m = 48°C$
$R^{H58}A$	R=AGA	A=GCT	
MR58A5	5' CTG ATT TAA **AGC** AGT TCT ATT 3'	21mer	$T_m = 44°C$
MR58A3	5' CC GAT AAT AGA ACT **GCT** TTA AAT 3'	23mer	$T_m = 48°C$

Table 4.2 Primary mutation oligonucleotides designed to mutate and replace the following residues: $R^{H50}A$, $R^{H56}A$, $R^{H58}A$ of the heavy variable chain and $Y^{L91}A$, $Y^{L94}A$, $Y^{L96}A$ of the light variable chain. The naming scheme of the primer is as follows: **M** Prefix - Mutation oligonucleotide, **5** suffix = 5' fragment, **3** suffix = 3' fragment. The underlined bases represent the codons encoding the residues to be mutated. (The MR58A3 displayed is the modified primer designed in response the dimerisation).

$D^{H54}L$	D=GAC L=CTG	
MD54L55'	GGA CAT CCG **GAC** TAT CAG GAC 3'	21mer $T_m = 60°C$
MD54L35'	GTC CTG ATA **CTG** CGG ATG TCC 3'	21mer $T_m = 60°C$
$H^{H52}A$	H=CAC A=GCT	
MH52aA5	5' ATC GGA AGG **AGC** AAT CCT GCC 3'	21mer $T_m = 66°C$
MH52aA3	5' GGC AGG ATT **GCT** CCT TCC GAT 3'	21mer $T_m = 66°C$
$Y^{L28}A$	Y=TAT A=GCT	
MY28A5	5' AGT GTT GCC **AGC** ACT ATG CAG 3'	21mer $T_m = 64°C$
MY28A3	5' CTG CAT AGT **GCT** GGC AAC ACT 3'	21mer $T_m = 64°C$
$Y^{L32}A$	Y=TAT A=GCT	
MY32A5	5' CCA ATA CAA **AGC** AGT GTT GCC 3'	21mer $T_m = 64°C$
MY32A3	5' GGC AAC ACT **GCT** TTG TAT TGG 3'	21mer $T_m = 64°C$
$Y^{L34}A$	Y=TAT A=GCT	
MY34A5	5' CAG GAA CCA **AGC** CAA TGA AGT 3'	21mer $T_m = 62°C$
MY34A3	5' ACT TAC TTG **GCT** TGG TTC CTG 3'	21mer $T_m = 62°C$
$Y^{L49}A$	Y=TAT A=GCT	
MY49A5	5' GGA CAT CCG **AGC** TAT CAG GAC 3'	21mer $T_m = 66°C$
MY49A3	5' GTC CTG ATA **GCT** CGG ATG TCC 3'	21mer $T_m = 66°C$

Table 4.3 Secondary mutation oligonucleotides designed to mutate and replace the following residues: $D^{H54}L$, $H^{H52}A$ of the heavy variable chain and $Y^{L28}A$, $Y^{L32}A$, $Y^{L34}A$, $Y^{L49}A$ of the light variable chain.

4.3 PRODUCTION OF OVERLAPPING PCR FRAGMENTS

Two oligonucleotides, LS1 and LS2, (Appendix) were designed to bind upstream and downstream of the insert on the vectors pKK233-2 and pTrc99A enabling PCR amplification of the 20G9 insert. These oligonucleotides were used in combination with the mutation oligonucleotides to enable production of the overlapping fragments for assembly of the mutant insert by overlap extension (Figure 4.6).

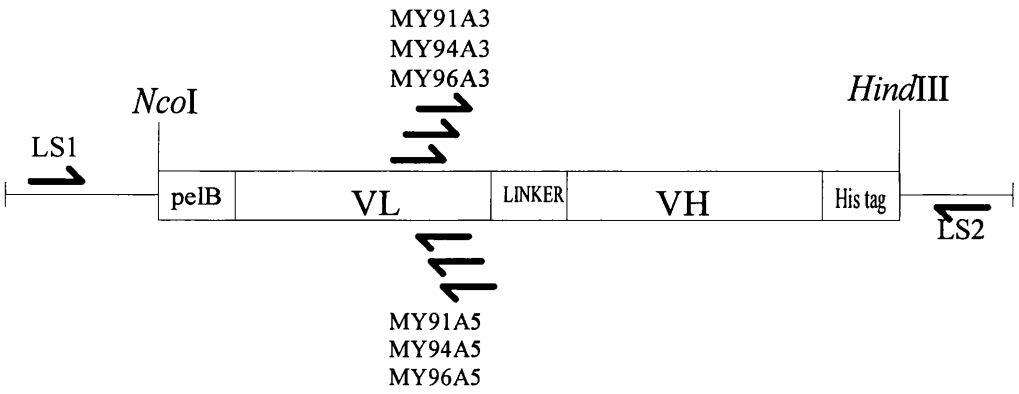


Figure 4.4 Mutagenesis of $Y^{L91}A$, $Y^{L94}A$, $Y^{L96}A$ within the L3 loop of the V_L. LS1 was used in combination with each upstream fragment oligonucleotide while LS2 was used in combination with each downstream fragment oligonucleotide.

4.3.1 Materials and Methods

4.3.1.1 Production of PCR fragments for assembly of the mutant inserts, $Y^{L91}{}_A$, $Y^{L94}{}_A$, $Y^{L96}{}_A$

5µl of Vent buffer and 5µl of the pre-mixed dNTP's (300µM) were added to eight 0.5ml thermo-tubes. 2µl of each 50pmol/µl oligonucleotide primer was added to the appropriate tube (Table 4.4) along with 1µl of the template DNA (0.25µg/µl)(excluding the -ve control), pKKpel/WT20G9 (LiCl plasmid preparation) and 0.5µl of BSA (10mg/ml). Water was added to bring the final volume to 49µl. 0.5µl of Vent DNA polymerase (2U/µl) was added and a drop of mineral oil was added to each tube. The mixtures were thermally cycled, 94°C x 1', 50°C x 1', 72°C x 40 secs) x 28. After the cycle was complete, the tubes were removed and stored at 4°C until required.

Table 4.4 1° mutations on V_L $Y^{L91}A$, $Y^{L94}A$, $Y^{L96}A$: PCR to produce overlapping fragments for the 1° mutations of $Y^{L91}A$, $Y^{L94}A$, $Y^{L96}A$. PCR conditions: 96°C 3 mins., [96°C 1min., 50°C 1min, 72°C 40 secs] x 28 cycles, 72°C 1min.

(ul)	1	2	3	4	5	6	-ve	+ve
Buffer	5	5	5	5	5	5	5	5
dNTP (300µM)	5	5	5	5	5	5	5	5
Oligo1(2ul)	MY91A5	MY91A3	MY94A5	MY94A5	MY96A5	MY96A5	MY91A5	LS1
Oligo2(2ul)	LS1	LS2	LS1	LS2	LS1	LS2	LS1	LS2
DNA (0.25ug/ul)	1	1	1	1	1	1	0	1
BSA (10mg/ml)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Vent (2U/ml)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
MilliQ	36	36	36	36	36	36	37	36
Tot. Vol.	50	50	50	50	50	50	50	50

Each fragment was resolved by electrophoresis on a 1% LMP agarose gel and the fragments excised and recovered using β-agarase (see 2.2.7). Purified fragments were visualised on a 1% agarose to estimate their relative quantities before addition to the PCR overlap extension reaction.

4.3.1.2 Production of PCR fragments for assembly of the mutant inserts, $R^{H}50A$, $R^{H}56A$ and $R^{H}58A$

10µl of Vent buffer and 10µl of the pre-mixed dNTP's (300µM) were added to eight 0.5ml thermo-tubes. 2µl of each 50pmol/µl oligonucleotide was added to the appropriate tube (Table 4.5) along with 0.25µl of the template DNA (LiCl plasmid preparation) (excluding the -ve control), pKKpel/WT20G9 and 1µl of BSA (10mg/ml). MilliQ water was added to bring the final volume to 49µl. 1µl of Vent DNA polymerase (2U/ml) was added and a drop of mineral oil was added to each tube. The mixtures were thermally cycled, 94°C x 1', 42°C x 1', 72°C x 40 secs) x 28. The exact details of the PCR are displayed in Table 4.5. After the cycle was complete, the tubes were removed and stored at 4°C until required.

Table 4.5 1° mutations on V_H $R^{H50}A$, $R^{H56}A$, $R^{H58}A$. PCR to produce overlapping fragments for the 1o mutations $R^{H50}A$, $R^{H56}A$, $R^{H50}A$. PCR conditions: 96°C 3 mins., [96°C 1min., 42°C 1min, 72°C 40 secs] x 28 cycles, 72°C 1min.

(ul)	-ve	+ve	+ve (#2)	1	2	3	4	5	6
Buffer	10	10	10	10	10	10	10	10	10
dNTP (300µM)	10	10	10	10	10	10	10	10	10
Oligo1 (2ul)	LS1	LS1	LS1	LS1	LS1	LS1	LS2	LS2	LS2
Oligo2 (2ul)	LS2	LS2	LS2	MR50A5	MR56A5	MR58A5	MR50A3	MR56A3	MR58A3
DNA (0.25µg/ml)	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
BSA (10mg/ml)	1	1	1	1	1	1	1	1	1
Vent (2U/ml)	1	1	1	1	1	1	1	1	1
MilliQ	74	73.75	73.75	73.75	73.75	73.75	73.75	73.75	73.75
Tot. Vol.	100	100	100	100	100	100	100	100	100

The PCR products were resolved by electrophoresing on a 1% LMP agarose gel. Each fragment was excised using longwave UV and recovered using β-agarase digestion, (section 2.2.7). The purified fragments were visualised and their relative quantities estimated by electrophoresing upon a 1% agarose against known quantities of 1Kb molecular weight markers, prior to their addition to the PCR overlap extension reaction.

4.3.2 Results

The production of six overlapping fragments for the subsequent overlap extension phase was semi-successful, producing fragments sizes of approximately 400-600 base pairs for the tyrosine residue mutations (Figure 4.7). However, the production of six overlapping fragments for use in the arginine overlap extension stage resulted in only five correctly sized fragments (Figure 4.8). This PCR was repeated several times unsuccessfully and so consequently the fragment $M^{H58}3$ was unavailable for the overlap extension phase thus forcing the abandonment of $R^{H58}A$ mutant production.

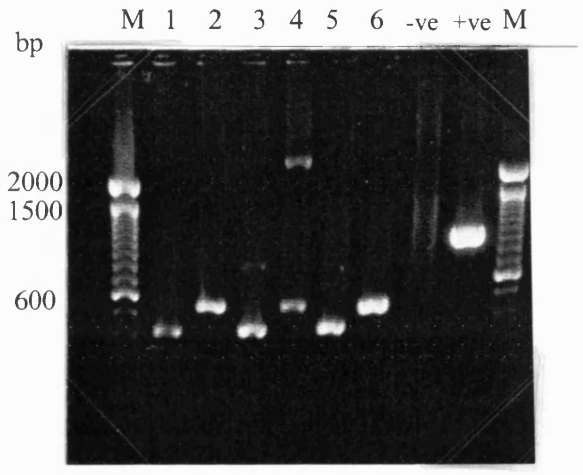


Figure 4.5 PCR production of overlapping fragments for the assembly of single mutation inserts resolved by electrophoresis on 1% agarose. 1) ~420bp $M^{L91}5$ fragment 2) ~580bp $M^{L91}3$ fragment 3) ~430bp $M^{L94}5$ fragment 4) 570bp $M^{L94}3$ fragment 5) ~440bp $M^{L96}5$ fragment 6) 560bp $M^{L96}3$ fragment, -ve: contains no template DNA, +ve: contains the full 20G9scFv WT construct DNA. M: 100bp ladder. *[Fragment sizes as predicted by DNA Amplify].*

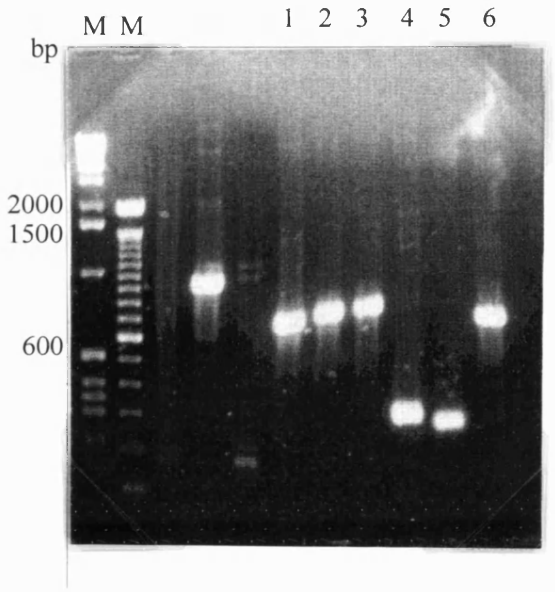


Figure 4.6 PCR production of overlapping fragments for the assembly of single mutation inserts resolved by electrophoresis on 1% agarose. [-ve] [+ve] [+ve 2*]. 1) ~660bp $M^{H50}5$ fragment 2) ~680bp $M^{H56}5$ fragment 3) ~690bp $M^{H58}5$ fragment 4) 310bp $M^{H50}3$ fragment 5) ~290bp $M^{H56}3$ fragment 6) (285bp)oversized $M^{H58}3$ fragment. Markers are 1Kb and 100bp ladders. *+*ve 2 - Promega kit plasmid preparation.* *[Fragment sizes are within parenthesis and were predicted using DNA Amplify].*

4.4 ASSEMBLY OF PCR FRAGMENTS

The overlapping PCR products when mixed together in PCR will extend to form a single fragment, in this case the 20G9 insert. Any alterations occuring in previous PCRs should be seen in the whole insert.

4.4.1 Materials and Methods

4.4.1.1 PCR overlap extension of Y^L91A, Y^L94A and Y^L96A fragments

5µl of Vent buffer and 5µl of the pre-mixed dNTP's (300µM) were added to five 0.5ml thermo-tubes. 0.5µl of each of the 50pmol/µl oligonucleotide primers, LS1 and LS2 were added to each tube (see Table 4.6) along with an appropriate amount of each of the two PCR fragments produced in 4.3.1.1 to act as assembly product. The -ve control excluded one PCR fragment demonstrating absence of contaminating template DNA). 0.5µl of BSA (10mg/ml) and water was added to bring the final volume to 49µl. 0.5µl of Vent DNA polymerase was added followed by a single drop of mineral oil to each tube. The mixtures were thermally cycled, (94°C x 30 secs., 50°C x 2', 72°C x 1') x 28, 72°C x 1'. The exact details of the PCR are displayed in Table 4.6.

Table 4.6 Overlap extension of the Y91A5/3, Y91A5/3, Y91A5/3 fragments. Conditions: 94°C 3 mins., [94°C 30 secs., 50°C 2mins., 72°C 1min.] x 28 cycles, 72°C 1min.

(ul)	1	2	3	-ve	+ve
Buffer	5	5	5	5	5
dNTP (300µM)	5	5	5	5	5
Oligonucleotide#1 (0.5ul)	LS1	LS1	LS1	LS1	LS1
Oligonucleotide#2 (0.5ul)	LS2	LS2	LS2	LS2	LS2
Frag. #1 (ul)	91A5 (5)	94A5 (5)	96A5 (3)	91A5 (2)	pKK20G9 DNA (1ul)
Frag. #2 (ul)	91A3 (5)	94A3 (5)	96A3 (5)	0	0
BSA (10mg/ml)	0.5	0.5	0.5	0.5	0.5
Vent pol. (2U/ml)	0.5	0.5	0.5	0.5	0.5
MilliQ water	28	28	30	36	37
Tot. Vol.	50	50	50	50	50

Removal of primers and PCR buffer was performed by buffer exchanging the PCR products into sterile water using Microspin S-200 columns followed by storage of the products at -20°C. Quantification of the inserts was determined at A_{260}.

4.4.1.2 PCR overlap extension of R^H50A and R^H56A fragments

The single PCR fragment for R58A was unavailable due to repeated failure of a PCR step. The reactions for R50A and R56A were performed twice with slight variations in the amount of template/extension material added. 5µl of Vent buffer and 5µl of the pre-mixed dNTP's were added to five 0.5ml thermo-tubes. 1µl of each of the 50pmol/µl oligonucleotide primers, LS1 and LS2, were added to the appropriate tube (see Table 4.7) along with an appropriate amount of the PCR product from 4.3.1.2 as assembly product along with 0.5µl of BSA (100mg/ml). Water was added to bring the final volume to 49µl. 0.5µl of Vent DNA polymerase was added followed with a drop of mineral oil and the mixtures thermally cycled, 94°C x 30secs., 50°C x 2', 72°C x 1') x 28, 72°C x 1'. The exact details of the PCR are displayed in Table 4.7. After the cycle was complete, removal of primers and PCR buffer was performed by buffer exchanging the PCR products into sterile water using Microspin S-200 columns, followed by storage of the products at -20°C. Quantification of the inserts was determined at A_{260}.

Table 4.7 Overlap extension of the R^H50A 3' & 5' fragments and the R^H56A 3' & 5' fragments. Conditions: 96°C 3 mins., [96°C 30 secs., 50°C 2mins., 72°C 1min.] x 28 cycles, 72°C 1min. The volumes of each fragment added are included within parenthesises).

(ul)	-ve	1	2	3	4
Buffer	5	5	5	5	5
dNTP	5	5	5	5	5
Oligonucleotide#1 (1ul)	LS1	LS1	LS1	LS1	LS1
Oligonucleotide #2 (1ul)	LS2	LS2	LS2	LS2	LS2
Frag. #1 (ul)	50A3 (2)	50A3 (4)	56A3 (5)	50A3 (2)	56A3 (3)
Frag. #2 (ul)	NA	50A5 (4)	56A5 (5)	50A5 (4)	56A5 (5)
BSA	0.5	0.5	0.5	0.5	0.5
Vent pol.	0.5	0.5	0.5	0.5	0.5
MilliQ water	35	33	27	35	33
Tot. Vol.	50	50	50	50	50

4.4.1.3 Sequencing of mutant inserts

100ng of each mutant insert was electrophoresed on a 1% LMP agarose gel and recovered from excised gel fragments using β-agarase digestion (section 2.2.6) and resuspended in sterile MilliQ water. The oligonucleotides *Link Seq. H* and *Link Seq. L* . (Appendix) were used for sequencing mutations in the heavy and light variable chains respectively. 3.2 pmol of oligonucleotide was added to a 0.2ml thin-walled PCR tube containing the 30-90ng of PCR product and the volume made up to 12μl with sterile MilliQ water. Each sample was sequenced on a 377 DNA sequencer ABI Prism (Perkin-Elmer) (2.2.10).

4.4.2 Results

4.4.2.1 Overlap extension

Overlap extension generated PCR products of approximately 850bp (Figure 4.9). Fragments of this size were expected and were clear evidence that successful overlap extension of the previous PCR products had occurred which allow amplification of the whole fragment using the two end primers, LS1 and LS2. The amount of template added to reactions 3 and 4 was estimated, based upon the fragment quantity after purification, to give a equimolar ratio of 1:1. This appeared to lead to a cleaner product than reactions 1 & 2 that were identical except for fragment quantity. The -ve control was missing a forward overlapping fragment preventing the formation of a whole insert, which subsequently would have been amplified by oligonucleotides LS1 & LS2.

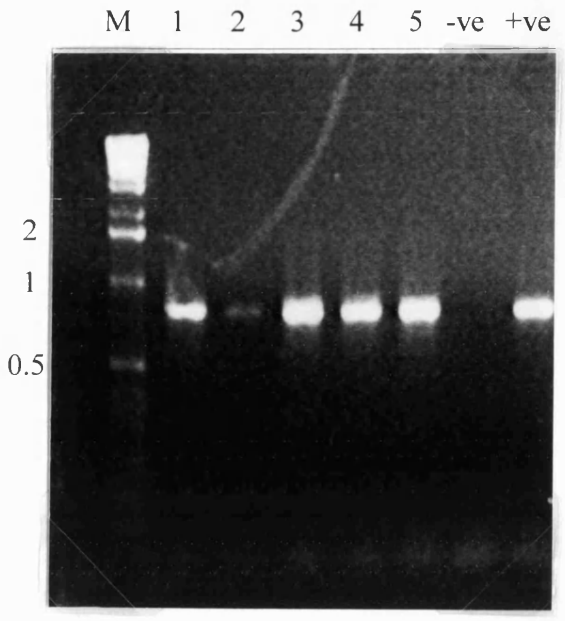


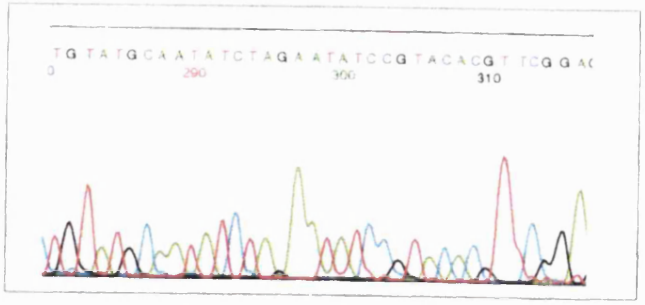
Figure 4.7 The assembly of the mutant inserts by overlap extension that produced equal sized fragments of approx. 970 bases. 1) $Y^{L}91A$ scFv, 2) $Y^{L}94A$ scFv, 3) $Y^{L}96A$ scFv, 4) $R^{H}50A$ scFv, 5) $R^{H}56A$ scFv, –ve) The -ve control lacked a forward overlapping fragment which prevented the formation of whole insert which subsequently would have been amplified by the oligonucleotides LS1 & LS2. +ve) plasmid template added. M) 1Kb ladder molecular weight marker. *The amount of template added to each reaction was estimated based upon the fragment quantity after purification to allow an equimolar ratio of 1:1.*

4.4.3 Sequencing of the PCR products

Each PCR product of the 20G9scFv insert containing each mutation was sequenced to verify the inclusion of the new residue and ensure that no spurious mutations had occurred elsewhere within the sequence. The chromatogram verified that all mutations were successful (Figures 4.10 and 4.11). The sequences of each mutation are compared to those of the wild type sequences in Table 4.9.

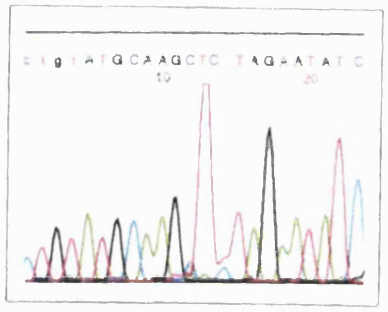
Table 4.8 Comparison of the DNA sequences of the 20G9 wild type DNA those of the mutant DNAs produced by overlap extension mutagenesis. In each case, substitution of 2-3 bases had occurred with the remaining sequence 5' and 3' to the mutated DNA shown to be error free.

20G9 WT:	-TGG ATT GGC AGG ATT CAT CCT-
Mutant R^H50A:	-TGG ATT GGC GCT ATT CAT CCT-
20G9 WT:	-TCC GAT AAT AGA ACT AGG TTA-
Mutant R^H56A:	-TCC GAT AAT GCT ACT AGG TTA-
20G9 WT:	-TGT ATG CAA TAT CTA GAA TAT-
Mutant Y^L91A:	-TGT ATG CAA GCT CTA GAA TAT-
20G9 WT:	-TAT CTA GAA TAT CCG TAC ACG-
Mutant Y^L94A:	-TAT CTA GAA GCT CCG TAC ACG-
20G9 WT:	-GAA TAT CCG TAC ACG TTC GGA-
Mutant Y^L96A:	-GAA TAT CCG GCC ACG TTC GGA-



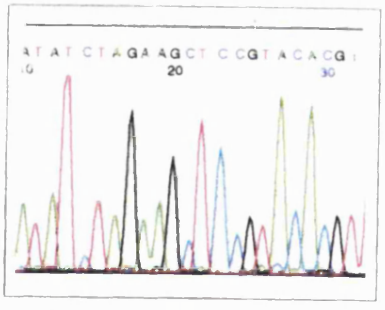
20G9 wild type sequence

-TGT ATG CAA **TAT** CTA GAA **TAT** CCG **TAC** ACG TTC GGA-
L91 L94 L96



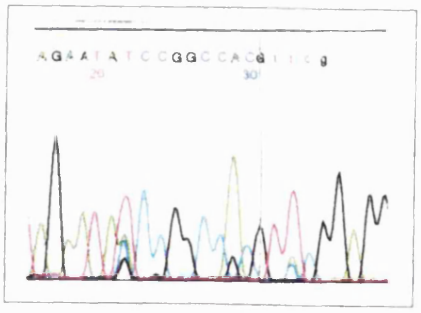
$Tyr^{L91}Ala$

-TGT ATG CAA **GCT** CTA GAA TAT-



$Tyr^{L94}Ala$

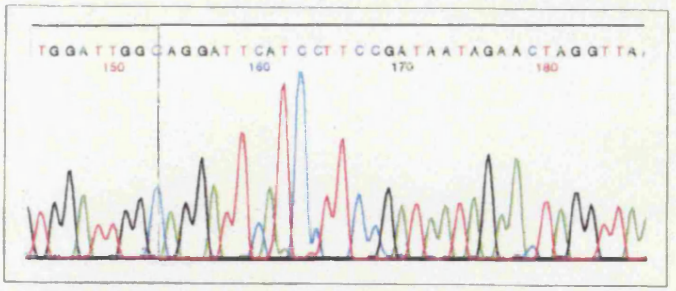
-TAT CTA GAA **GCT** CCG TAC ACG-



$Tyr^{L96}Ala$

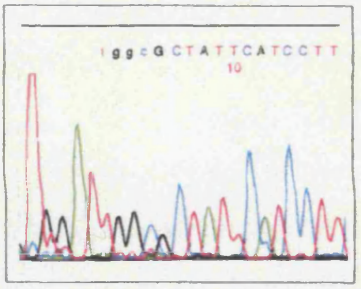
-GAA TAT CCG **GCC** ACG TTC GGA-

Figure 4.8 377 DNA sequencer ABI prism chromatogram of the tyrosine mutations, $Y^{L}91A$, $Y^{L}94A$, $Y^{L}96A$ sequences compared directly to the chromatogram sequence of the 20G9 wild type sequence.



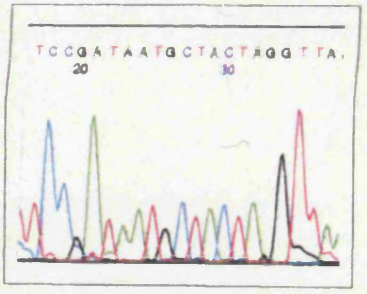
20G9 wild type sequence

-TGG ATT GGC **AGG** ATT CAT CCT TCC GAT AAT **AGA** ACT AGG TTA-
H50 H56



$Arg^{H50}Ala$

-TGG ATT GGC **GCT** ATT CAT CCT-



$Arg^{H56}Ala$

-TCC GAT AAT **GCT** ACT AGG TTA-

Figure 4.9 377 DNA sequencer ABI prism chromatogram of the arginine mutations, $R^{H}50A$, $R^{H}56A$ sequences compared directly to the chromatogram sequence of the 20G9 wild type sequence.

4.5 DISCUSSION

4.5.1 Design of oligonucleotides for mutagenesis

The site-directed mutagenesis was successful in producing residue changes at the positions discussed. No spurious mutations were seen at the mutagenic site or in any other part of the sequence. The extension overlap mutagenesis using Vent DNA polymerase did not introduce any erroneous mutations that are commonly reported when using Taq DNA polymerase. This is almost certainly due to the high fidelity of the Vent DNA polymerase.

Overlapping fragments were produced by PCR for the mutants $Y^{L}91A$, $Y^{L}94A$, $Y^{L}96A$ $R^{H}50A$ and $R^{H}56A$. Of the PCR products for the mutant $R^{H}58A$, only one of the two necessary fragments was of the correct size. This PCR reaction was repeated several times but repeatedly resulted in a product of ~600 bases. This could have been due to non-specific binding, incorrect synthesis of the MR58A3 oligonucleotide or due to poor binding of the oligonucleotide containing the shorter complementary strand within its 3' region. Consequently this mutation was abandoned. The mutants $Y^{L}91A$, $Y^{L}94A$, $Y^{L}96A$ $R^{H}50A$ and $R^{H}56A$ were all successfully extended by PCR overlap extension to fragments of approximately 850bp. These PCR fragments were all successfully sequenced using the primers Link Seq L, Link Seq H and NEB pUC119 sequencing primer. All confirmed that the planned mutation had occurred and verified that no erroneous mutations had been introduced elsewhere into the scFv fragment sequence by polymerase error.

Selected codons were checked against the optimal codon tables observed for expression in *E. Coli* [Ikemura, 1981]. However even if the codons used were inadvised, codon usage within the gene was given priority. Altering a single codon so that it was optimal for *E. coli* expression while the vast majority of surrounding codons could be considered poor would probably have been pointless.

CHAPTER 5

5 EXPRESSION OF 20G9 SCFV

5.1 INTRODUCTION

5.1.1 Expression of scFv

The advent of the single chain antibody (scFv) [Bird *et al.*, 1988, Huston *et al.*, 1988] was heralded as the solution to the antibody engineer's problems - a molecule that had the ability to bind like an antibody but at one fifth of the size. Its relatively small size made it potentially useful for therapeutics, particularly in cancer therapy [King *et al.*, 1994]. The scFv was a development of the Fv [Skerra & Pluckthun, 1988; Field *et al.*, 1988] in which the two variable domains are covalently joined using a flexible peptide linker, overcoming the problem of chain dissociation of the Fv at low concentrations. The scFv has been shown to fold in the same conformation as the variable regions of the parent antibody [Zdanov *et al.*, 1994] and demonstrates similar binding affinity to that of the whole antibody [Pantoliano *et al.*, 1991]. However, it is now apparent that the expression of scFv in *E. coli* can be extremely difficult, regardless of the stability of the whole parent antibody. This is often manifested in a low level of expression in stark contrast to the high levels of most other recombinant proteins expressed in *E. coli*. This may be due to the presence or absence of key residues that determine scFv stability in the expression host, and hence whether or not it will be favourably expressed [Knappik and Pluckthun, 1994].

The linkers that bridge the V_H and V_L were engineered to span a ~35Å gap between the carboxy terminus and amino terminus of the light and heavy chains, permitting correct folding but with no interference in binding. However the length and the sequence of the peptide linker is known to have an adverse effect on expression levels [Huston *et al.*, 1991; Deonarain *et al.*, 1997]. Most linkers are usually between 14-25 residues in length with the 15 residue $(Gly_4Ser)_3$ linker [Huston *et al.*, 1988], currently the most commonly used. Composition is important in that polar residues such as serine, glutamate, aspartate can enhance solubility by

allowing hydrogen bonding to solvent molecules, whereas the more simple glycine residues add flexibility to the linker preventing conformational disturbance of the Fv and its binding site [Argos, 1990].

There are two main methods of obtaining antibody fragments from *E. coli*. The first is to produce scFv within cytoplasmic inclusion bodies. High levels of protein can be expressed in the cytoplasm of *E.coli* where it accumulates as insoluble aggregates, [Huston *et al.*, 1988] forming between 5-10% of total cell protein [Bird *et al.*, 1988]. Once the cystosolic contents are extracted, the non-functional scFv requires *in vitro* folding using urea and/or guanidium hydrochloride. However, relatively low yields of active antibody may be obtained due incorrect folding of the scFv and be due to the incorrect formation of the intra-domain disulphide bridges [Proba *et al*, 1995; Buchner *et al.*, 1992].

The second method of expression is for the cell to secrete the antibody fragment into the periplasmic space. The incorporation of a bacterial signal peptide, such as the *pelB* gene (pectate lysase from *E. carotovora*) [Kustersvan-Someren *et al.*, 1992] or STII [Picken *et al.*, 1983] enables the antibody fragment to be transported into the periplasm. Secretion into the periplasmic, a more oxidative compartment than the cytoplasm, can assist folding of the scFv by the formation of the correct disulphide bridges [Glockshuber *et al.*, 1990]. Correct folding generally means that the antibody is soluble at low levels. In addition, antibody within the periplasm is localised and results in a relatively easy method from which to extract the active scFv, with the added benefit of containing very few contaminating proteins compared to in the cytoplasm. However, high level expression is thought to lead to aggregation of scFv within the periplasmic space [Kipriyanov *et al.*, 1994]. With such high levels of protein in the small periplasmic space, the formation of insoluble aggregates may be favoured over correct folding [Kiefhaber *et al.*, 1991]. Kipriyanov *et al.*, 1997 has found that the addition of sucrose to the growth medium leads to a increase in the yield of soluble scFv antibody. The sucrose is thought to prevent the formation of insoluble aggregates and allow soluble scFv to be passed into the growth supernatant from which it can be purified.

5.1.2 Peptide tags for detection and purification

Peptide tags are useful in allowing identification of the antibody fragment and detection at low levels, an important feature when optimising small scale expression. In addition tags can be used to purify the antibody fragment from the cell cytoplasm, the periplasm or the culture medium.

The tag used for detection of 20G9scFv and its mutants in this thesis is the c-myc tag, a peptide containing the epitope of the proto-oncogene c-myc and recognised by the monoclonal antibody 9E10 (kindly donated by M.W. Robertson, The Scripps Research Inst., La Jolla, USA) [Evan *et al.*, 1985]. The antibody can be used to immunoblot expressed tagged scFv. A positive blot generally indicates in-frame production of the protein, as the tag is at the C-terminus of the protein sequence. Immunoblotting is also extremely sensitive, so that low levels of protein can be detected. This allows protein expression on a small scale where conditions can be altered in order to optimise expression, with increased signal intensity indicating improved expression.

The second tag used on the 20G9 scFv was the histidine tag [Hochuli et al., 1988; Yip *et al.*, 1989]. This consists of a short repeating sequence of histidine residues on the C-terminus of the protein. The imidazole rings of the histidine residues complex with transition metal ions such as Ni^{2+}, Zn^{2+}, Fe^{2+} and Co^{2+} with a high affinity. If these ions are immobilised onto a support then proteins with many histidine residues become bound to the support via the ions while other proteins without this affinity will elute from the support. This is the basis of immobilised metal affinity chromatography (IMAC) [Porath *et al.*, 1975] (Figure 5.1). A benefit of IMAC is that relatively mild elution conditions only are needed to disrupt protein binding. Imidazole can be added to compete for metal coordination sites displacing the protein or strong chelators such as EDTA that strip the support of its metal ions can be used. A further method involves lowering the pH resulting in protonation of the histidine residues causing release of the bound protein.

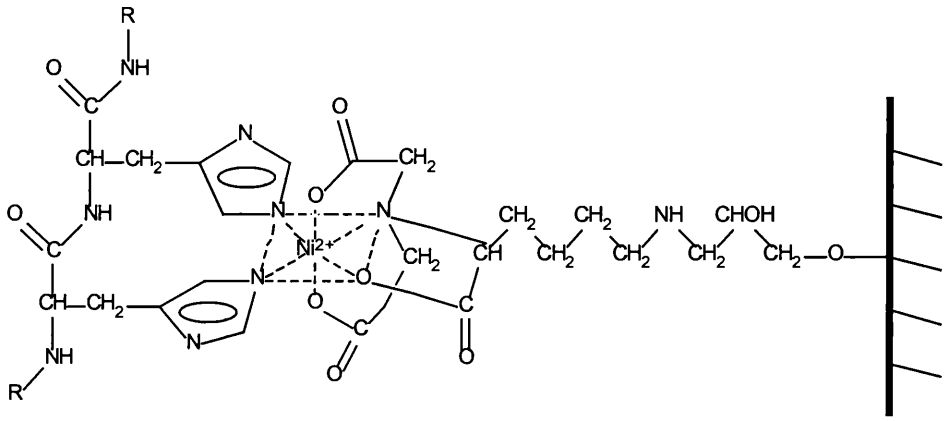


Figure 5.1 Interaction between a histidine tag and Ni-NTA resin (Qiagen). The basis of IMAC purification is that the imidazole rings of the histidine co-ordinate with the metal ions on the support which enable the tagged protein to bind while un-tagged proteins pass through during the washing pre-elution phase (picture reproduced from Qiagen literature).

5.1.3 A survey of suitable expression vectors: problems and solutions

5.1.3.1 pKK233-2

The esterolytic antibody 20G9 was cloned and assembled by PCR into the scFv format [Smith *et al.*, 1993]. The light and heavy variable chains were joined with the 14 residue peptide linker (EGKSSGSGSESKVD) in the unconventional (but unimportant) order of V_L-V_H [Whitlow & Filpula, 1991]. The scFv also included a C-terminal histidine tag of five residues in length. The antibody fragment was cloned into the expression vector pKK233-2 (Clontech Laboratories Inc., Palo Alto, USA; distributors for Pharmacia Biotech) which was the form in which it was supplied for the work in this thesis (Figure 5.2).

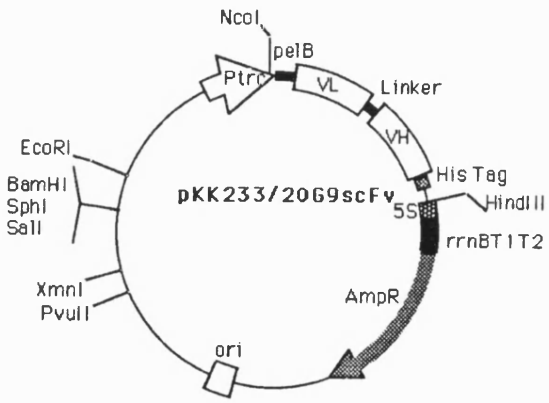


Figure 5.2 20G9 scFv was cloned into the vector pKK233-2 and the construct named pKK233/20G9scFv. A single base in the restriction site *Nco*I was deleted disabling the site and start codon within it, allowing the ATG codon of *pelB* to initiate 'in frame' translation.

The restriction sites into which the scFv was cloned were *Nco*I/*Hind*III, the first and last restriction sites within the multiple cloning site (MCS) of pKK233-2. The *Nco*I restriction site, CCATGG, encodes an ATG codon which initiates protein translation, a deliberate feature of this vector. If translation initiates at this *Nco*I start codon instead of the start codon of the the *pelB* signal peptide then the scFv is expressed 'out of frame'. This was prevented by a single base deletion of G from the ATG codon allowing translation to initiate at the ATG codon of the *pelB* signal peptide [Smith, 1994].

The pKK233/20G9scFv construct was supplied in the *E. coli* strain XL-1 Blue as a stab culture. Glycerol stocks of the culture were prepared and stored at -70°C along with plasmid preparations of the construct that were stored at -20°C and -70°C. An [1]IGEN expression protocol for scFv expression was used, growing the cells in 2xYT containing 100μg/ml ampicillin and 15μg/ml tetracycline. The culture was grown at 30°C to an A_{600} of 1.5 followed by induction with 1mM IPTG at 26°C for 3 hours from which protein was obtained through periplasmic lysis. However expression levels fell and poor yields were improved by changing to the *E.coli* strain TG-1, but again fell after initial success. The pKK233-2 vector was later withdrawn by Pharmacia Biotech almost certainly due to plasmid instability [Bell, 1994; Elliot, 1994], later known to have been a relatively common phenomenon when using this vector. A replacement vector, pTrc99A, was donated by Pharmacia Biotech.

5.1.3.2 pTrc99A

The prokaryotic expression vector pTrc99A (Pharmacia Biotech) was the succesor to the vector pKK233-2. pTrc99A (Figure 5.3) holds several improvements over its predecessor. It includes a *lacI*^{q} gene encoding the *lac* promotor repressor protein allowing strong repression before induction preventing pre-induction 'leakage' and a *trc* promotor, a hybrid of the *trp* and *tac* promotors and a stronger promoter than the *lac* and *tac* inducible promotors. The vector had also been designed with the promotor, the ribosome binding site (RBS) and the MCS in quick succession. The optimal positioning of these sites with respect to the start of the cloned gene is known to be a factor that strongly influences expression. The restriction site *Nco*I is the first site within the MCS and encodes an ATG codon eliminating the need for a start codon within any insert.

The *pel*B leader sequence upstream of the 20G9scFv gene encodes a start codon for translation initiation for 'in-frame' expression of the antibody fragment. To clone the 20G9 scFv gene into pTrc99A, the *Nco*I site must be used but the *pelB* start codon is a single base downstream of the start codon within the *Nco*I site. If the restriction site ATG codon is left intact, translation of the scFv is out of frame. Deletion of the bases between the two ATG codons can eliminate the restriction site

[1]IGEN International Ltd. are a biotech company based in Rockville, Maryland, USA that had a commercial interest in the early catalytic antibodies.

but may introduce alterations into the periplasmic leader sequence which is critical for functional secretion. Deletion of at least a single base of the *Nco*I ATG codon would be necessary to shift initiation of translation to the pelB ATG codon and allow expression of 20G9 scFv.

```
AGGAAACAGACC ATG GAA TTC GAG CTC GGT ACC CGG GGA TCC TCTAGA GTCGACCTGCAGGCATGCAAGCTT
             |     ||-------|      |     |-|---|-|     |       |       |     |     |    |
Ptrc         --------       ----------------    ----------------------------------------
      NcoI   EcoRI   SacI    KpnI  SmaI  BamHI  XbaI   SalI   PstI  SphI HindIII
```

pTrc99A multiple cloning site

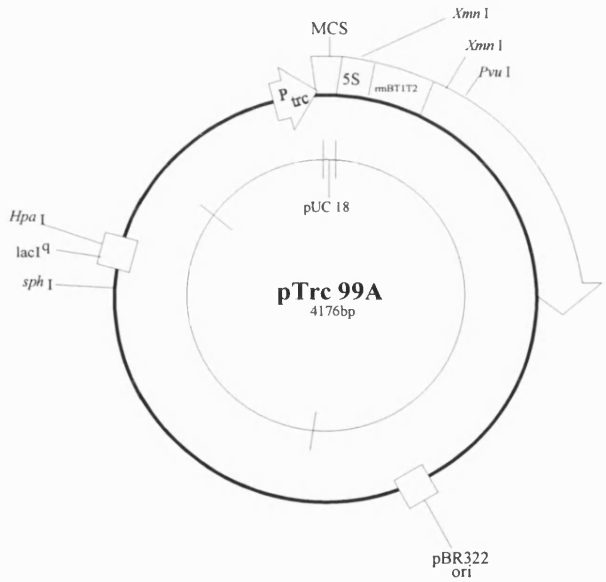
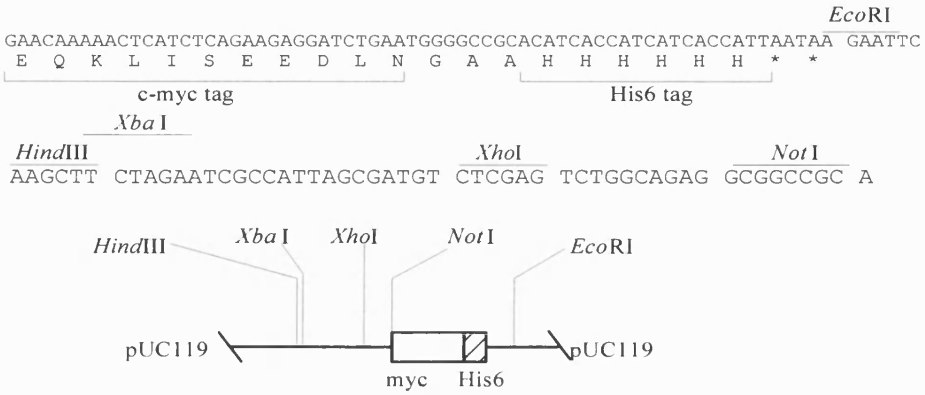


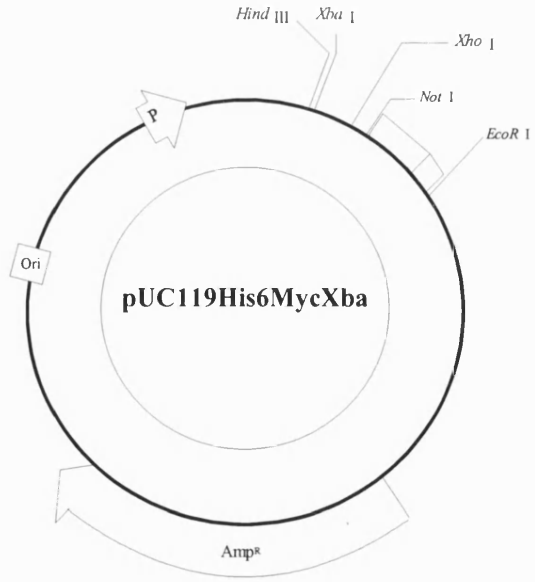
Figure 5.3 The expression vector pTrc99A uses the P*trc* promotor, a hybrid of the *trp* and *tac* promotors. The *lac*Iq repressor gene helps reduce uninduced expression. The restriction site *Nco*I is the first site within the MCS encoding an ATG codon eliminating the necessity for an ATG codon on a potential insert.

5.1.3.3 *pUC119His6mycXba*

The vector pUC119His6*mycXba* [Griffiths *et al.*, 1994] (kindly donated by G. Winter, MRC, Cambridge, UK) is derived from the pUC119 vector [Vieira and Messing, 1987] and contains four restriction sites of which the *Hind*III site is closest to the promoter (Figure 5.4). Downstream of the *Not*I site is a *c-myc* and His6 tag that can be used for detection and purification. The vector promotor is *lac* and the vector confers ampicillin resistance. 20G9 scFv did not encode these restriction sites and so engineering of the scFv DNA was necessary to introduce the relevant restriction sites. The penta-His tail of 20G9 scFv required removal by PCR due to the existence of the *myc* and histidine tags in the vector.



i)



ii)

Figure 5.4 i) sequence of the pUC119HisMycXba multiple cloning site. ii) the vector pUC119His6*mycXba* uses the *lac* promotor and includes the *c-myc* and a hexa-histidine tag downstream of the MCS. The vector confers ampicillin resistance to the cell.

5.2 CLONING INTO THE VECTOR: PTRC99A

5.2.1 Methods

The *Nco*I restriction site of pKK233/20G9scFv was previously disabled by a single deletion. Therefore it was first necessary to re-engineer the site into 20G9scFv enabling it to be easily restricted and ligated into the pTrc99A vector. Disabling the ATG codon within the *NcoI* restriction site after construction required that the site was restricted using *NcoI* followed by exposure to mung-bean nuclease, specifically degrading single-stranded DNA (ie.. sticky-ends of restricted DNA). However it was important to remove only the single stranded DNA as further degradation of proximal DNA would have resulted in the deletion of the ATG codon of the pelB leader sequence, which was critical for successful expression. Only EndA$^-$ *E. coli* strains, i.e. XL-1 Blue or DH5α, were used due to reduced nuclease activity on the DNA preparations.

5.2.1.1 Engineering the NcoI restriction site into 20G9scFv by PCR

1µl of pKK233/20G9scFv plasmid (LiCl plasmid preparation, 250ng/ml) was mixed with 5µl Vent buffer, 5µl of mixed dNTPs and 0.5µl BSA (100ug/ml). 0.5µl of the oligonucleotide primers NCOITRC (100ng/ul) and LS2 (100ng/ul) (Appendix) were added and the volume raised to 49.5µl with water. Finally 0.5µl of Vent DNA polymerase (2U/ul) was added followed by a drop of mineral oil (2.2.8). The control was water substituted for DNA. Each reaction was thermally cycled (94°C x 1', 55°C x 1', 72°C x 1') x 28, 72°C x 1' and the PCR product passed through a MicroSpin S-200 column (Pharmacia Biotech) to buffer exchange and remove primers (2.2.5).

5.2.1.2 Restriction of 20G9 scFv and the vector pTrc99A

A restriction enzyme digestion (2.2.4) was performed on 10µl (200ng/µλ) of the PCR product and 0.5µl (200ng/ml) of the vector pTrc99A. 2µl of NEBuffer 2 and 0.5µl of *Nco*I (10U/µl) and *Hind*III (20U/µl) were added to both and each volume taken to 20µl. Three controls were included where only a single or no restriction enzyme was added. The reactions were incubated at 37°C overnight followed by

heat inactivation. The DNA samples were each run on a LMP agarose gel and the restricted insert and digested vector removed and recovered using β-agarase digestion (2.2.6) and quantified by UV at A_{280}.

5.2.1.3 Ligation of 20G9 scFv into the vector pTrc99A

4µl of the restricted 20G9scFv insert and 1µl of the restricted pTrc99A vector were mixed with 1µl of ligation buffer, 0.5 T4 DNA ligase (20U/µl) and the volume taken to 10µl with water. The ligation control substituted water for the insert. The ligation reaction was incubated at 16°C overnight followed by heat inactivation and buffer exchanged in MicroSpin columns (2.2.6). The ligations were transformed into *E.coli* strain TG1 and DNA minipreps (2.3.1) prepared from subsequent colonies.

5.2.1.4 Exonuclease reduction of pTrcpel/WT20G9 using Mung-Bean nuclease

A restriction enzyme digestion (2.2.4) was performed on 3µl (200ng/µl) of the pTrcpel/WT20G9 construct to which 2µl of NEBuffer 2 and 1µl of *Nco*I (10U/µl) where added. The volume was taken to 20µl and incubated at 37°C overnight followed by heat inactivation. The control was substitution of restriction enzyme by water. Restricted construct was buffer exchanged in a MicroSpin column (Pharmacia Biotech). Mung-Bean nuclease (10U/µl) (NEB) was diluted 1:9 in the supplied mung-bean nuclease buffer (NEB) (~0.6U per 1µg DNA). In duplicate, 2µl of the restricted pTrcpel/WT20G9 construct (500ng/µl) were mixed with 0.6µl of nuclease, 2.5µl of 10x buffer and the volume taken to 25µl with water. The control was water substituted for nuclease. The reactions were heated at 30°C for 10 and 30 minutes and inactivated by the addition of 0.01% SDS. The DNA was recovered by ethanol precipitation at -20°C overnight and resuspended in 5µl of Millipore water. 2µl of each were run on 1% agarose to verify that the DNA remained essentially intact.

5.2.1.5 Blunt-end ligation of the pTrcpel/WT20G9 construct

1µl of the blunt-ended pTrcpel/WT20G9 construct and the nuclease negative controls were mixed with 1µl of ligation buffer (NEB), 0.5 T4 DNA ligase (20U/µl)(NEB) and the volume taken to 10µl with water. The ligation reaction was incubated at 16°C overnight followed by heat inactivation. The ligation reactions were buffer exchanged in MicroSpin columns (2.2.6) and transformed into *E. coli* strain TG1. 5 colonies from each of the two plates were selected which possessed the putative *Nco*I^{-ve} pTrcpel/WT20G9 construct and DNA minipreps performed (2.3.1). 2µl of each DNA miniprep was restricted with 0.5µl *NcoI* to screen for constructs containing disabled *NcoI* sites. Those failing to restrict were sequenced to enable examination of the disabled site and its surrounding DNA.

5.2.2 Results

5.2.2.1 Cloning 20G9 scFv into the pTrc Expression Vector

The PCR produced inserts of approximately 850bp in length which were restricted with *NcoI* and *HindIII* along with the pTrc99A vector and purified and buffer exchanged. After ligation, transformation of the constructs produced several hundreds of plaques of which 10 were PCR screened for the 20G9 insert from which 9 were found to be positive.

5.2.2.2 Disabling the NcoI site of the pTrcpel/WT20G9

The two positive clones after restriction and exposure to mung-bean nuclease had approximately the expected length. The blunt ended linear plasmid was ligated and transformation of the ligation mix produced approximately 150 colonies. Of the 10 selected 50% failed to digest when exposed to the restriction enzyme *Nco*I which was indicative of a disabled site (Figure 5.5).

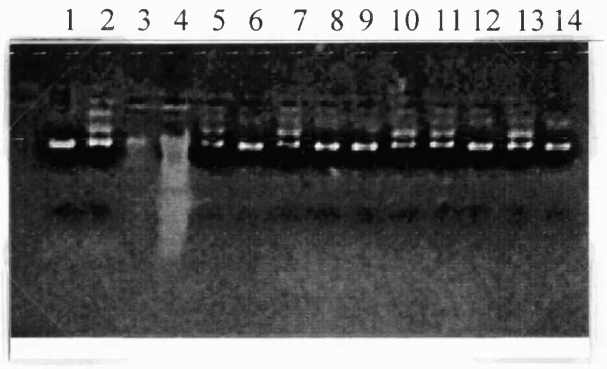


Figure 5.5 Screening for *Nco*I disabled construct. 1) +ve restriction control 2) -ve restriction control 3) M.Wt. Marker 4) M.Wt. Marker. Lanes 5-14 contain selected constructs of which 5, 7, 10, 11 and 13 fail to restrict with *Nco*I which indicated that this restriction site had been successfully disabled.

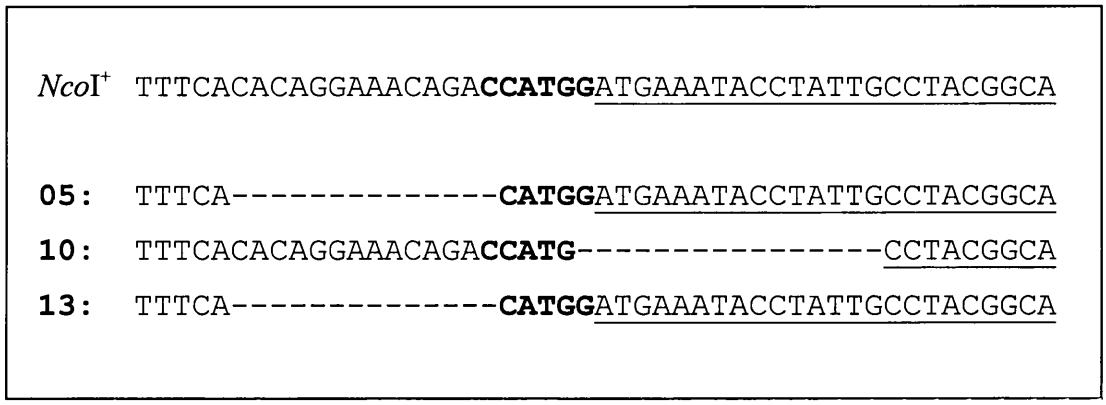


Figure 5.6 The flanking DNA sequences of the disabled *Nco*I site of three nucleased pTrcpel/WT20G9 constructs. Construct 5 and 13 both deleted 3'-5' disrupting the site but leaving the ATG codon of the site intact. Construct 10 a disabled *Nco*I but also had many bases of the pelB leader sequence deleted.

Sequencing of three of the constructs allowed examination of the nuclease-treated region (Figure 5.6). It was found that the nuclease degraded the single-stranded DNA but repeatedly degraded double-stranded DNA also. In the above examples the ATG codon of the *Nco*I site was either left intact or the nuclease degraded the pelB leader sequence. The example shown was the best of four attempts that included different conditions (20°C, 30°C and 5, 10 and 30 minutes). However, these resulted in either a single nuclease digestion or resulted in large deletions of the scFv gene. Only after lengthy repeated failures involving restriction, nuclease digestion, blunt-end ligation, screening and sequencing, was the experiment abandoned and a decision taken to attempt cloning into a new vector, pUC119His6*mycXba*.

5.3 CLONING INTO THE VECTOR: pUC119His6*mycXba*

5.3.1 Methods

5.3.1.1 Engineering the HindIII and NotI restriction sites into 20G9scFv

Two oligonucleotide primers were designed (HINDFOR & NOTBACK, Appendix) that introduce a *Hind*III and *Not*I restriction site 5'and 3' respectively of 20G9scFv. 1µl of pKK233/20G9scFv plasmid was mixed with 5µl Vent buffer, 5µl of mixed dNTPs and 0.5µl BSA (100ug/ml). 0.5µl of the oligonucleotide primers HINDFOR

and NOTBACK were added and the volume taken to 49.5µl with water. Finally 0.5µl of Vent DNA polymerase was added followed by a drop of mineral oil (2.2.8). Controls were used by substituting DNA for water. The mixture was thermally cycled (94°C x 1', 50°C x 1', 72°C x 1') x 28 and the PCR product passed through a MicroSpin column to buffer exchange and remove primers (2.2.5).

5.3.1.2 Cloning the 20G9scFv insert into pUC119His6mycXba

A restriction enzyme digestion (2.2.4) was performed upon 10µl (200ng/µl) of the PCR product and 0.5µl (200ng/ml) of the vector pUC119His6*mycXba*. 2µl of NEBuffer 3 and 0.5µl of *Not*I (10U/µl) and *Hind*III (20U/µl) were added to the vector and 1.5µl of *Not*I and *Hind*III added to the insert and the volumes taken to 20µl with water. Three controls were included where only a single enzyme or no restriction enzyme was added to the vector. The reactions were incubated at 37°C overnight followed by heat inactivation. The DNA samples were each run on a LMP agarose gel and the restricted insert and vector recovered using β-agarase digestion (2.2.6) and quantified by UV at A_{280}.

5.3.1.3 Verification of insert digestion

Insertion of the restricted 20G9scFv into the restricted pUC119His6*mycXba* vector proved difficult. In order to verify that the PCR insert was digesting with the restriction enzymes the 5' termini of the insert were labelled with ^{32}P (2.2.11) and the insert restricted with *Hind*III and *Not*I to confirm that the ends of the PCR fragment had been removed.

12µl of the restricted 20G9scFv insert and 3µl of the restricted pUC119His6*mycXba* vector were mixed with 1µl of ligation buffer, 0.5 T4 DNA ligase (20U/µl) and the volume taken to 10µl with water. A ligation control was water substituted for insert. The ligation reactions were incubated at 16°C overnight followed by heat inactivation and buffer exchange in MicroSpin columns (Pharmacia Biotech.) (2.2.6). Each ligation mixture was transformed (2.3.5) into the *E.coli* strain XL-1

Blue and the transformed cells plated onto LB plates (2% (w/v) glucose, 100μg/ml ampicillin, 15μg/ml tetracycline) and incubated overnight at 37°C. Subsequently, 15 colonies of at least 2mm diameter were picked from the plate and screened for recombinant plasmid using protoplast screening (2.3.6).

5.3.1.4 Small scale expression of scFv

Two freshly transformed plaques were used to inoculate 250ml flasks containing 50ml of XL media and appropriate antibiotics. The cultures were then incubated at 30°C with shaking at 250RPM until reaching a density of A_{600} 1.5-2.0. The cells were then induced with 1mM IPTG and incubated at 26°C with shaking at 250RPM for the specified time. The cells were harvested at 6000 x g using a Sepatech Megafuge 1.0 benchtop centrifuge (Heraeus, Brentwood, UK). The pellets were resuspended in 10 ml of osmotic shock buffer (30mM Tris/HCl pH 8.0, 20% (w/v) sucrose, 1mM EDTA), incubated at room temperature with shaking for 10 minutes and pelleted at 6000 x g, (10ml buffer was added per 50ml of culture). The supernatant was discarded and the pellet resuspend in 20ml of 5mM $MgSO_4$ and incubated at 4°C with shaking for 10 minutes. Finally the lysed cells were spun at 8500 x g and the supernatant lysate collected. Any large particulates were removed from the lysate by passing through a 0.45μM filter (Millipore). All lysates were then concentrated in a Centricon-10 (Amicon) to a final volume of 0.5ml. The concentrated lysates were loaded onto a 12% SDS-PAGE gel and examined for scFv with Coomasie blue stain or by Western blotting targetted to the *myc* tag.

5.3.1.5 Western Blotting of the tagged 20G9 scFv

15μl of each of the two concentrated cell lysates were loaded onto 12% SDS-PAGE (2.2.13) along with the positive control of F_{ab} possessing a *myc* tag and low molecular weight rainbow markers (Amersham International, Amersham, UK). The gel was blotted onto nitrocellulose membrane (2.2.14) and after blocking with reconstituted powdered milk the membrane was incubated with the primary monoclonal antibody, 9E10 (donated by M. Robertson, The Scripps Research Institute, La Jolla, USA) overnight at 4°C. The membrane was washed and

incubated for 2 hours at room temperature with the secondary polyclonal antibody, horse radish peroxidase (HRP) conjugated to goat anti-mouse antibody. The antigen-antibody complex was located by immersion of the membrane in developing solution (2.2.14), fixed in water, dried and stored in darkness to prevent fading.

5.3.1.6 Sequencing of the constructs

Constructs positively located by Western blotting were sequenced (2.2.10) to verify that no PCR induced mutations had been introduced to the scFv. Two sequencing primers were used (Appendix): LACB which annealed to the construct upstream of the insert allowing sequencing of the pelB, V_L and linker, LSH annealed to the 20G9scFv linker sequence and allowed sequencing of the V_H and myc and His tags.

5.3.2 Results

5.3.2.1 Cloning 20G9 scFv into pUC119His6mycXba

A PCR product of approximately 850bp in size was obtained, as determined by the molecular weight ladders which corresponded with the values calculated using DNA Amplify. The insert and vector were digested with *Hind*III and *Not*I (Figure 5.7), with digestion of the 20G9scFv insert verified by radiolabelling the ends of the insert followed by its restriction (Figure 5.8).

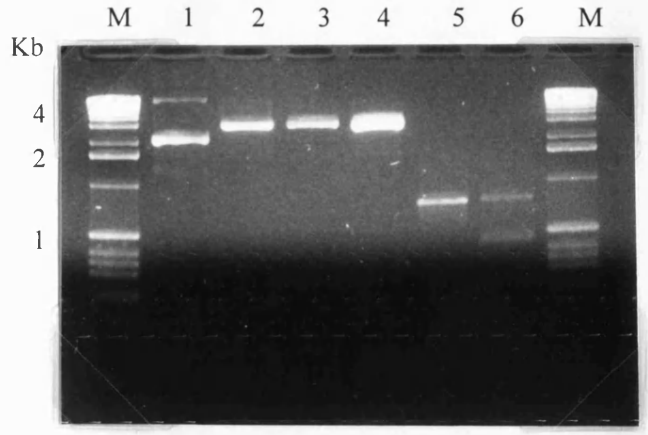


Figure 5.7 Restriction of the pUC119His6*mycXba* vector and the 20G9scFv insert by *Hind*III and *Not*I enzymes. 1) unrestricted vector 2) *Not*I restriction of vector 3) *Hind*III restriction of vector 4) *Not*I/*Hind*III double restriction of vector 5) *Not*I/*Hind*III double restriction of insert 6) unrestricted insert.

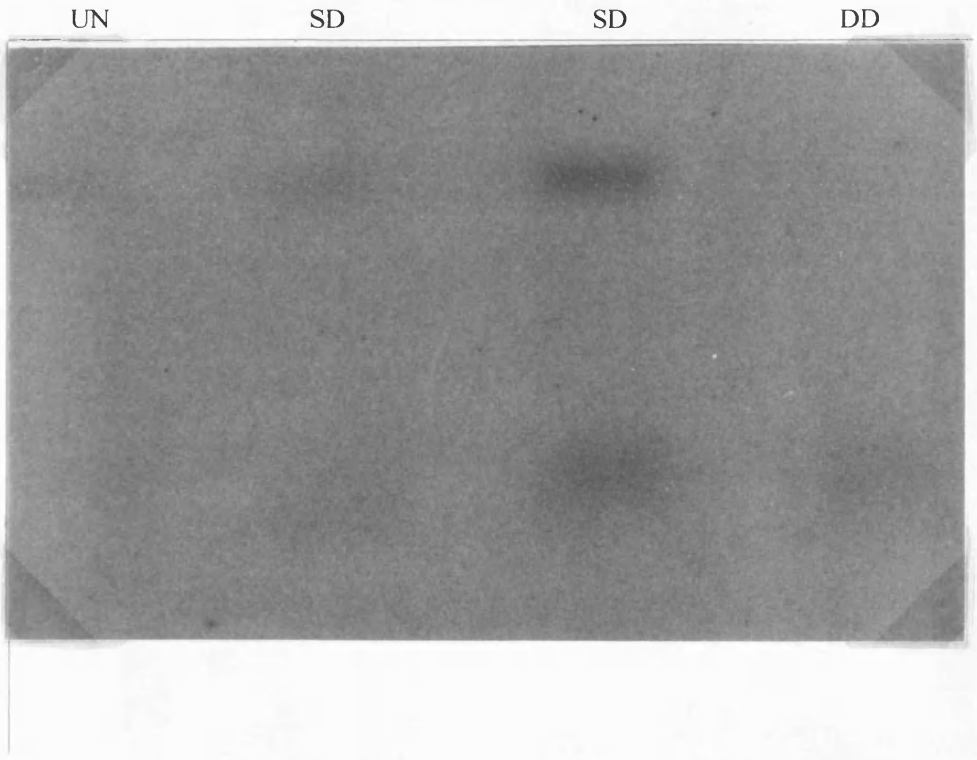


Figure 5.8 The 20G9scFv insert ends were radiolabelled with ^{32}P and digested with *Not*I and *Hind*III. (1) no restriction (2) *Hind*III only (3) *Not*I only (4) double digestion with *Hind*III and *Not*I. This verifies restriction of the inserts when the levels are too low to visualise by UV on agarose gels.

5.3.2.2 Protoplast screening for the pUC-20G9-His6myc construct

Two recombinants were identified from the 15 plaques screened (Figure 5.9).

5.3.2.3 Expression and Western blot detection of 20G9scFv

20G9scFv was detected in the cell lysate by Western blotting with 9E10 (Figure 5.10). The scFv was approximately 31.5kDa in size determined by the molecular weight markers. In addition the approximate mass of the protein was calculated based upon the average molecular weight of amino acids.

5.3.2.4 Sequencing of the constructs

SNAP mini-preps prepared from the two positive expressors denoted pUC-20G9-His6*myc* [H] and [10] were sequenced with LACB and LSH (Appendix) and revealed no mutation of either sequence (the full DNA sequence for pUC-20G9-His6*myc* is shown in Appendix).

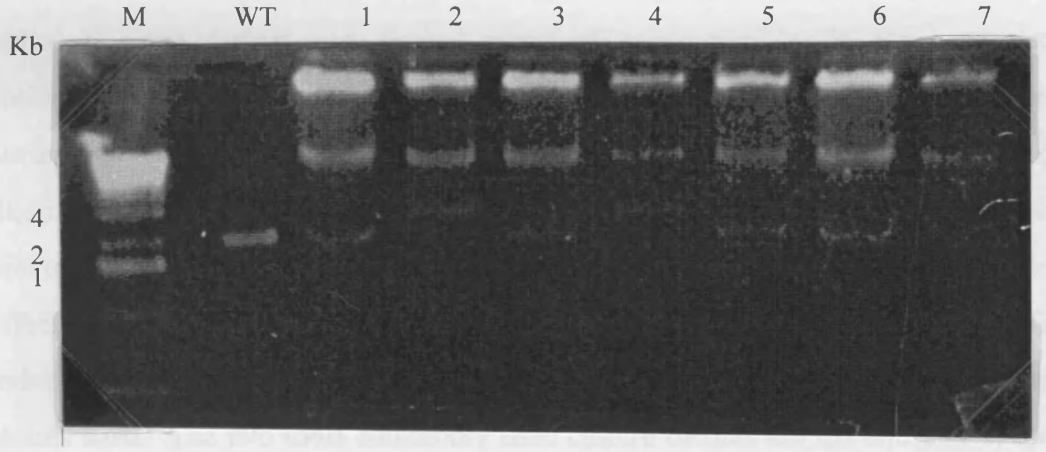


Figure 5.9 Protoplast screening for recombinants after ligation of wild type 20G9scFv into the vector pUC119His6*mycXba*. WT control (wild-type vector without insert) 1-7) Screened plaques. M - 1Kb ladder. 2 and 4 display recombinants which are of a higher molecular weight than the WT control. The second row (not displayed) contained no recombinants.

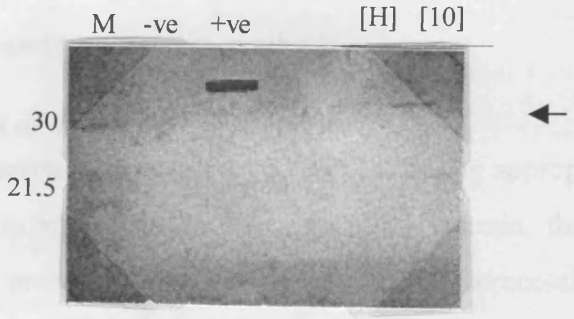


Figure 5.10 12% SDS-PAGE gel of cell lysate of the constructs pUC-20G9- His6*myc* [H] and [10] analysed by Western blotting with the antibody 9E10. Protein of 31.5kDa in size possessing the *Myc*-tag is detected for both constructs.

5.4 EXPRESSION AND PURIFICATION OF 20G9 SCFV

Once cloned into the final expression vector the conditions for the expression of 20G9 scFv were optimised. There exist typical expression conditions for scFvs, although some factors can favour increased expression levels for particular antibodies. A factor common to many expression protocols are the temperatures during the growth (30°C) and induction (26°C) phases [Somerville *et al.*, 1994; Skerra and Pluckthun, 1991], adherence to which have been shown to improve yields of functional protein. However, many other variables can also have dramatic affects on yields: the media volume within the conical flask, the amount of IPTG inducer added, the length of induction in addition to the culture media and bacterial strains used. The two most commonly used culture medias are LB and 2xYT, the latter simply twice as concentrated as the first. However, Griffiths *et al.*, 1994 demonstrated that increasing the relative levels of yeast extract to tryptone improved the levels of scFv expression. This novel growth media was named XL media (15g trypton, 15g yeast extract, 10g NaCl for 1L) and was used in the expression of 20G9scFv and its mutants. *E. coli* strains used for expression of 20G9scFv were TG-1 and TOPP2.

5.4.1 Expression and purification methods

5.4.1.1 Optimisation of the 20G9scFv expression

Several small scale expressions (15ml XL media containing appropriate antibiotics) were performed and examined the bacterial expression strain, the induction time and the necessity of protease inhibition in downstream processing. The *E.coli* strains TG1 and TOPP2 were compared during induction times of 1, 3, 5, 8 and 24 hours. The effect of the protease inhibitor PMSF was also assessed.

5.4.1.2 Large scale expression of scFv

A freshly transformed colony was used to inoculate a starter culture of 20ml of XL media containing appropriate antibiotic. The culture was incubated at 30°C in a shaker incubator (250 RPM) until A_{600} = 0.8. From this culture, 1 ml was used to inoculate a 2L flask containing 500ml XL media containing appropriate antibiotic. The culture was allowed to incubated at 30°C in the shaker incubator overnight. Fresh antibiotic and IPTG was added to a final concentration of 1mM and the incubator temperature reduced to 26°C and the culture allowed to induce for 24 hours. Cells were harvested by centrifugation at 5,000 x g for 15 minutes at room temperature and the pellet resuspended in osmotic shock buffer (30mM Tris/HCl pH 8.0, 20% sucrose, 1mM EDTA). 1ml of buffer was added per 5ml culture, followed by incubation at room temperature with vigorous shaking for 10 minutes. The solution was centrifuged at 8,000 x g for 10 minutes at 4°C after which the supernatant was discarded. The pellet was resuspended in ice cold 5mM $MgSO_4$ by adding an amount equal to that of the osmotic shock buffer. The mixture was incubated at 4°C with vigorous shaking followed by centrifugation at 8,000 x g for 10 minutes at 4°C. The supernatant lysate was decanted and filtered through 0.45μM filters (Millipore) and stored at 5°C.

5.4.1.3 N'-terminal His-Tag purification

300μl of Ni^{2+} charged NTA-resin slurry (Qiagen) was added to 100ml of filtered cell lysate and incubated with shaking at 4°C for 1 hour. The resin was allowed to settle for 30 minutes and carefully decanted into a 5ml Econo-column (Bio-Rad,

Hemel Hempstead, UK). Any remaining resin was retrieved by gentle centrifugation (50 x g) of the lysate from which the resin was recovered and added to the column. The resin was washed with wash buffer (50mM NaPO4 pH 8.0, 300mM NaCl, 20mM imidazole) using a peristaltic pump until the A_{260} of the breakthrough was identical to the wash buffer (approx. 250ml). The column was connected to a Biologic LP low pressure chromatography system (Bio-Rad) and the scFv eluted from the column by an Imidazole gradient using the elution buffer (50mM NaPO4 pH 8.0, 300mM NaCl, 250mM imidazole) with the correct peak collected in 1ml aliquots. The absorbance of each aliquot was measured at A_{280} in order to verify the location of scFv and the aliquots pooled appropriately. The scFv was concentrated and buffer exchanged with assay buffer (10mM Tris pH 8.8, 140mM NaCl) in a Centricon-10 membrane filter (Amicon). The concentration of the protein was measured using a BCA protein assay kit (Pierce) and diluted to approximately 250µg/ml with assay buffer. If the scFv preparation was not used immediately, it was snap frozen using an acetone and dry ice bath and stored at -70°C until futher use.

5.5 RESULTS

5.5.1 Small scale expression and purification

5.5.1.1 Optimising the expression and secretion conditions

1mM IPTG was found to be optimal for induction of 20G9scFv. *E.coli* strain TOPP2 led to increased yields of scFv in comparison to *E.coli* strain TG-1. Increased yields of periplasmically secreted scFv were seen after 3 and 24 hours induction at 26°C. The level of scFv within the cell supernatant was highest after 24 hours while the addition of PMSF appeared to have no observable effect (Figures 5.11-5.14).

5.5.1.2 N'-terminal His-Tag purification

Periplasmic lysate and the cell supernatant both contained 20G9scFv. The scFv was easily purified from the periplasmic lysate by batch binding with Ni^{2+} charged NTA-resin (Figure 5.15) but it was not possible to purify the scFv from large volumes of cell supernatant.

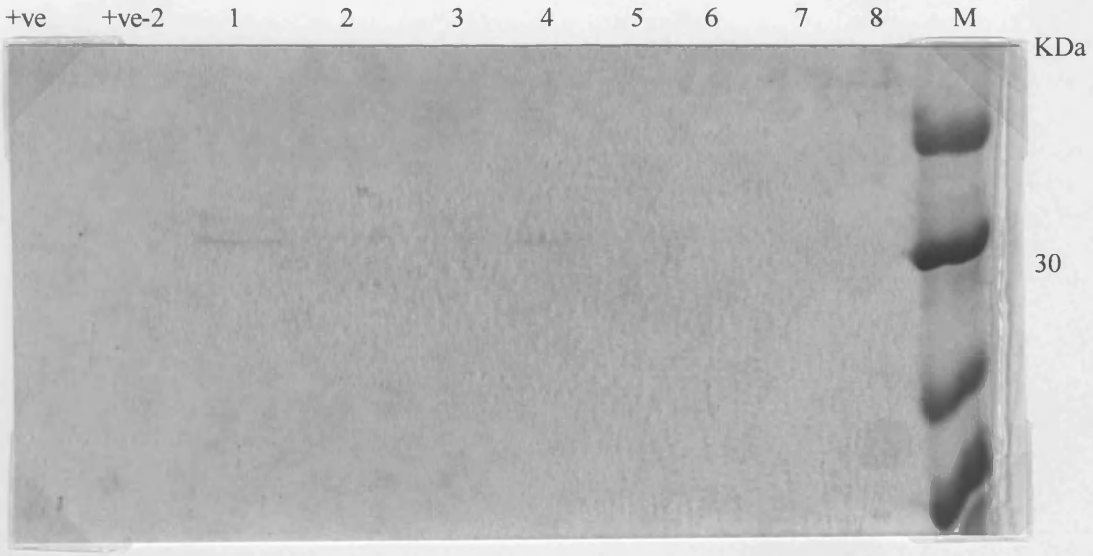


Figure 5.11 Comparison of 3hr vs 24hr IPTG induction periods with 20G9scFv expression in *E.coli* strains TOPP2 and TG1. The addition of PMSF protease inhibitor was assessed in the same study. +/- = with/withoutPMSF. 1) TOPP2, 3hr, - 2) TOPP2, 3hr, + 3) TOPP2, 24hr, - 4) TOPP2, 24hr, + 5) TG1, 3hr, - 6) TG1, 3hr, + 7) TG1, 24hr, - 8) TG1, 24hr, +. The IMAC purified protein was resolved by 12% SDS-PAGE and stained with Coomasie blue.

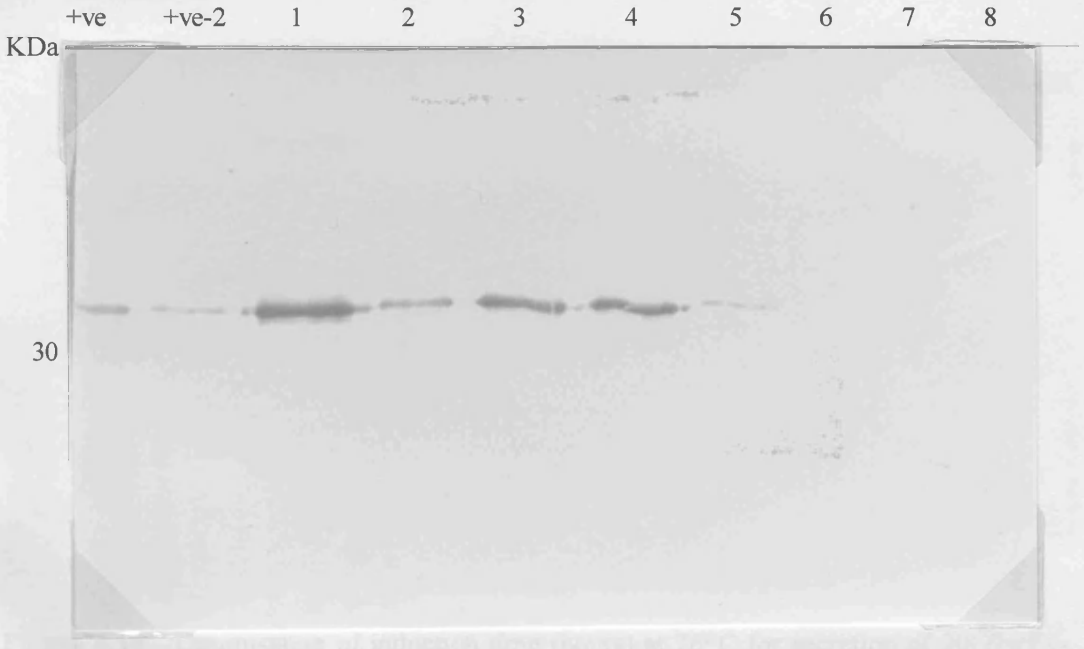


Figure 5.12 Comparison of 3hr vs 24hr IPTG induction periods of 20G9scFv expression in *E.coli* strains TOPP2 and TG1 and identified by Western blotting with 9E10. The addition of PMSF protease inhibitor was assessed in the same study. +/- = with/withoutPMSF. 1) TOPP2, 3hr, - 2) TOPP2, 3hr, + 3) TOPP2, 24hr, - 4) TOPP2, 24hr, + 5) TG1, 3hr, - 6) TG1, 3hr, + 7) TG1, 24hr, - 8) TG1, 24hr, +. The IMAC purified protein was resolved by 12% SDS-PAGE and stained with Coomasie blue.

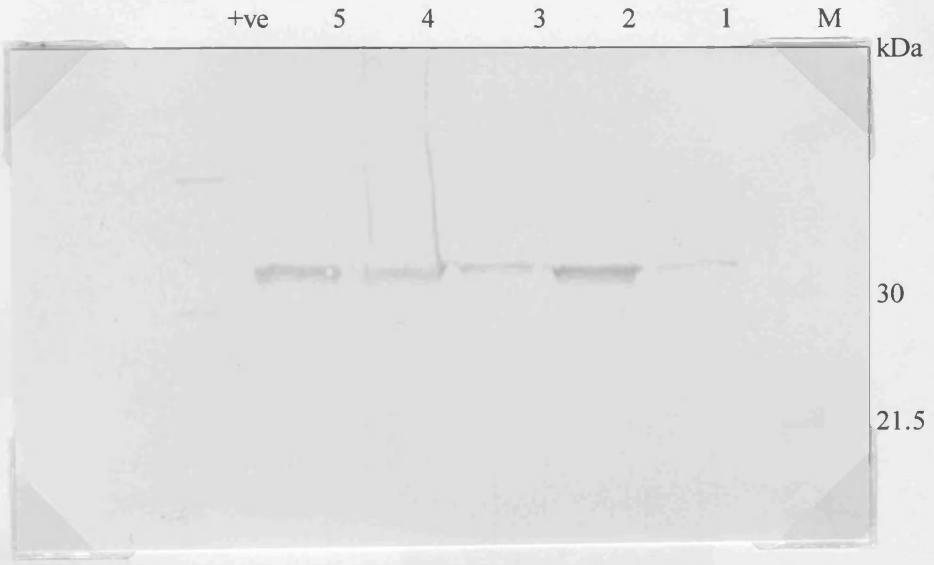


Figure 5.13 Optimisation of induction time (hours) at 26°C for periplasmic secretion of 20G9scFv-WT in *E.coli* strain TOPP2. Each cell lysate was resolved on 12% SDS-PAGE and analysed by Western blotting using 9E10 anti-*myc* tag antiserum. 1) 2hr 2) 3hr 3) 5hr 4) 8hr 5) 24hr [+ve: myc-tagged Fab].

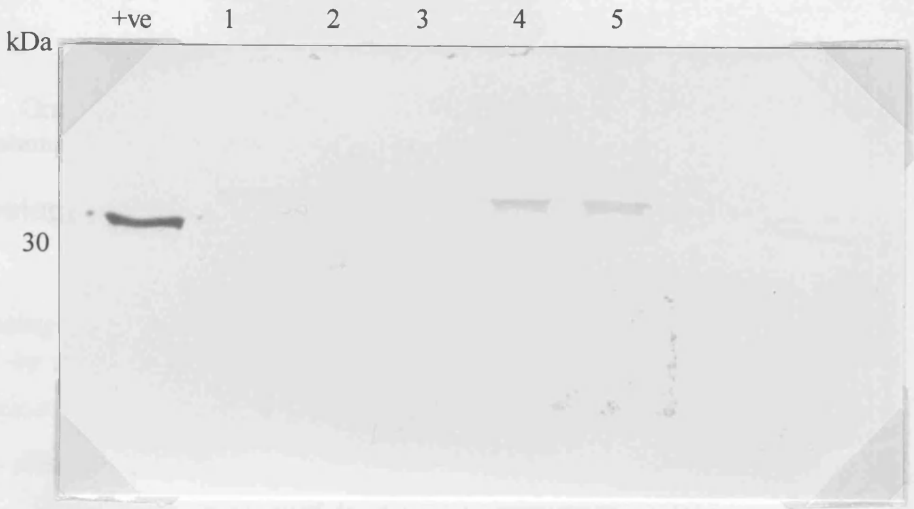


Figure 5.14 Optimisation of induction time (hours) at 26°C for secretion of 20G9scFv-WT into the supernatant from *E.coli* strain TOPP2. Each supernatant was first concentrated by membrane centrifugation and resolved on 12% SDS-PAGE. Analysis was by Western blotting using 9E10 anti-*myc* tag antiserum. [+ve: *myc*-tagged Fab], 1) 2hr 2) 3hr 3) 5hr 4) 8hr 5) 24hr.

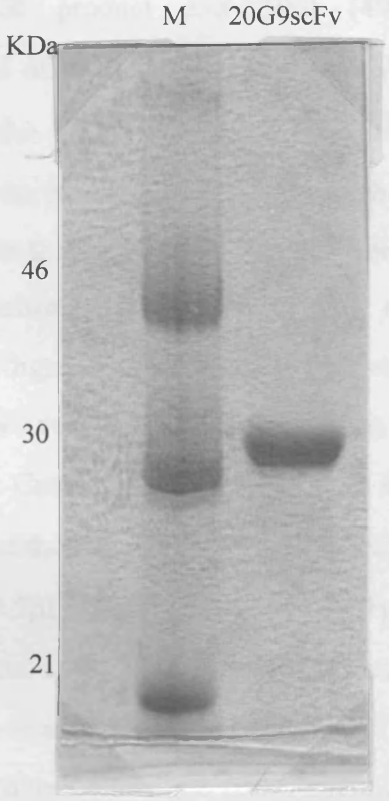


Figure 5.15 Gradient IMAC purified 20G9scFv resolved by 12% SDS-PAGE and visualised by staining with Coomasie blue R-250.

5.5.2 Cloning, expression and purification methods to produce scFv mutants.

5.5.2.1 Cloning of the 20G9scFv mutants, Y^L91A, Y^L94A, Y^L96A, R^H50A, R^H56A, into the vector pUC119His6mycXba.

The oligonucleotide primers HINDFOR & NOTBACK, (Appendix) were used to introduce the *Hind*III and *Not*I restriction sites 5'and 3' respectively onto the 20G9 mutant PCR products. 1µl of PCR product for Y^L91A, Y^L96A, R^H50A, R^H56A and 3µl of the Y^L94A PCR product were mixed with 5µl Vent buffer, 5µl of mixed dNTPs and 0.5µl BSA (100ug/ml). 0.5µl of the oligonucleotide primers HINDFOR and NOTBACK were added and the volume taken to 49.5µl with water. Finally 0.5µl of Vent DNA polymerase was added followed by a drop of mineral oil (2.2.8). The negative control was water substituted for template DNA and the positive

control was pKK233/20G9scFv added as the template. The mixture was thermally cycled (94°C x 30 secs., 50°C x 2', 72°C x 1') x 28 and the PCR products passed through a MicroSpin column (Pharmacia Biotech) to buffer exchange and remove the primers (2.2.6). A restriction enzyme digestion (2.2.4) was performed on 45µl (200ng/ml) of the PCR product and 10µl (400ng/ml) of the vector pUC119His6*mycXba*. 6µl of NEBuffer 3 and 2µl of *Not*I (10U/µl) and *Hind*III (20U/µl) were added to the PCR products and 1µl of *Not*I and *Hind*III to the pUC119His6*mycXba* and the volumes taken to 60µl with water. The reactions were incubated at 37°C overnight followed by heat inactivation. The DNA samples were each run on a LMP agarose gel and the restricted mutant inserts and vector recovered using β-agarase digestion (2.2.7) and quantified by UV at A_{280}.

The restricted inserts and vector were ligated using a Fast-Link ligation kit (Epicentre Technologies, Cambridge, UK) (2.2.5). 0.5µl of restricted vector (10U/µl) and 7µl of each of the restricted inserts were mixed with 1.5µl of Fast-link ligation buffer (10x) and 1.5µl ATP solution(10mM). 1µl of the T4 DNA ligase (2 U/µl) was added, the volume taken to 15µl with water and the reaction incubated at room temperature for 1 hour and heat inactivated at 70°C for 15 minutes. 2.5µl of each ligation mixture was transformed into *E.coli* strain XL-1 Blue and plated onto LB media containing 2%(w/v) glucose, ampicillin (100µg/ml), tetracycline (15µg/ml) and incubated in darkness at 37°C overnight. 10 colonies were selected from each plate, except that containing $Y^{L}94A$ where only 5 colonies were selected, and screened for recombinant plasmid using protoplast screening (2.3.6). A DNA SNAP mini-prep (2.3.1) was performed on the two colonies that screened positive for the mutants and the plasmid preparations sequenced (2.2.10).

5.5.2.2 Expression and purification of the 20G9scFv mutants

Each mutant scFv was first expressed on a small scale based on the method in section 5.2.1.2 and *myc*-tagged antibody fragments were immunodetected using the antibody 9E10 (2.2.14). Positively detected mutant scFvs were expressed on a large scale (5.4.1.2), IMAC purified (5.4.1.3), snap frozen and stored at -70°C if not used immediately.

5.5.3 Results

5.5.3.1 Cloning of the 20G9scFv mutants, Y^L91A, Y^L94A, Y^L96A, R^H50A, R^H56A, into the vector pUC119His6mycXba.

All the mutants Y^L91A, Y^L94A, Y^L96A R^H50A and R^H56A that had been successfully extended by PCR overlap extension in section 4.4.2.1 had *Hind*III and *Not*I restriction sites introducted by PCR. The inserts were successfully restricted (Figure 5.16 & 5.17) and ligated into the pUC119His6*mycXba* vector. All 10 colonies were positive for the constructs pUC-Y^L91A-His6*myc*, pUC-R^H50A-His6*myc* and pUC-R^H56A-His6*myc* while 7 out of 10 were positive for the pUC-Y^L96A-His6*myc* constructs (Figures 5.18 & 5.19). Two from five colonies were positive for the pUC-Y^L94A-His6*myc* construct. Sequencing of one of each of the mutant constructs showed no PCR induced errors by Vent polymerase.

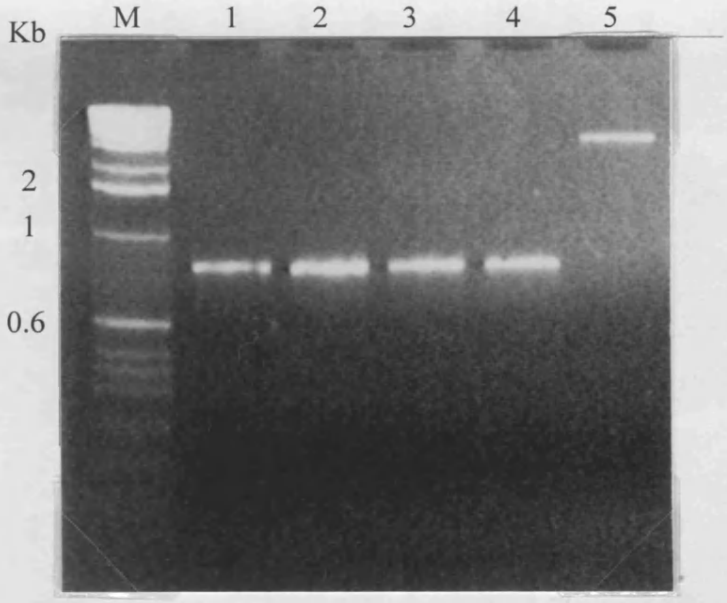


Figure 5.16 PCR products and the pUC119His6*mycXba* vector after restriction with the *Hind*III and *Not*I enzymes: M) 1 Kb ladder 1) Y^L91A 2) Y^L96A 3) R^H50A 4) R^H56A 5) restricted pUC119His6*mycXba*. Each insert had the restriction sites *Hind*III and *Not*I introduced by PCR. However the PCR involving the Y^L94A mutant gave a low yield of product and was repeated separately (see below).

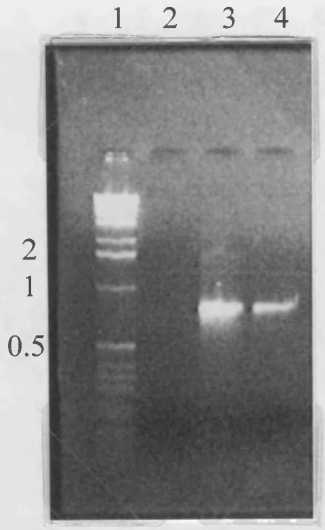


Figure 5.17 The PCR product of the $Y^{L}94A$ mutant after an additional round of PCR 1) 1 Kb ladder 2) -ve control 3) $Y^{L}91A$ 4) +ve control.

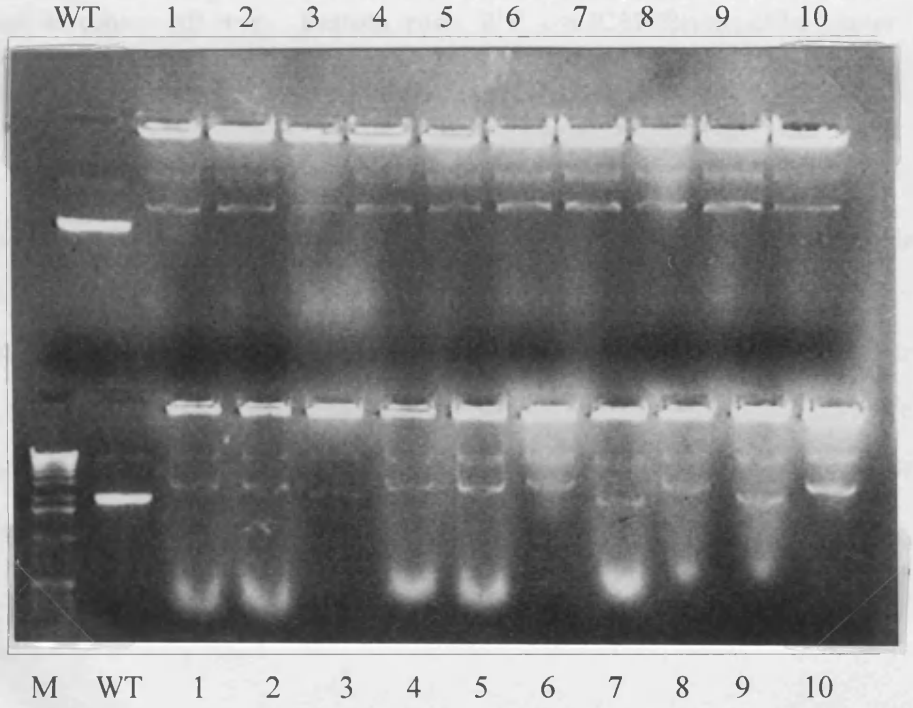


Figure 5.18 Protoplast screening for the $Y^{L}91A$ & $Y^{L}96A$ constructs based upon increased molecular weight. **Top row**: WT- pUC119His6*mycXba* vector, 1-10) $Y^{L}91A$ protoplast screens. All +ve. **Bottom row**: WT - pUC119His6*mycXba* vector 1-10) $Y^{L}96A$ protoplast screens. (5,9 & 11 -ve only).

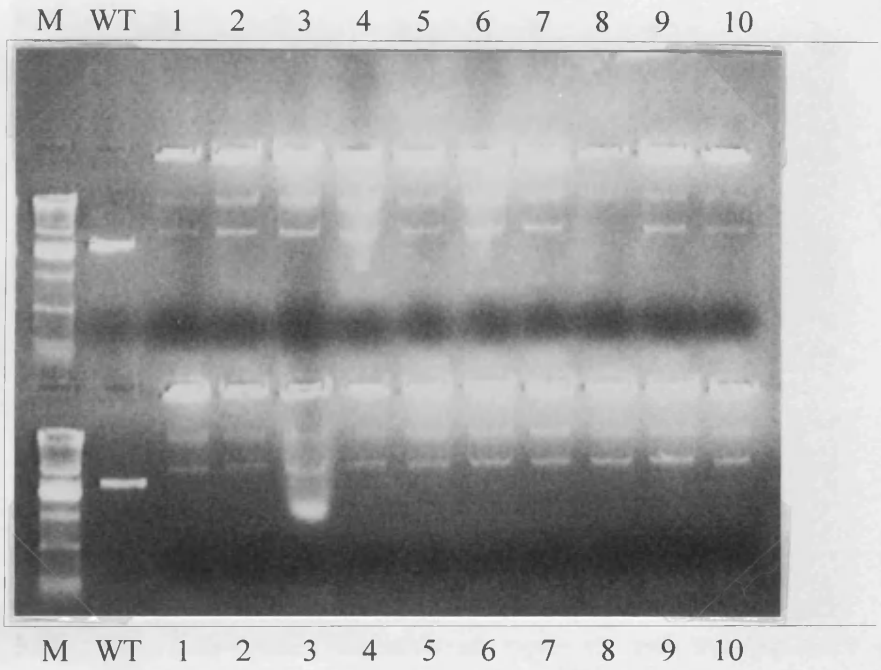


Figure 5.19 Protoplast screening for the $R^{H}50A$ & $R^{H}56A$ constructs based upon increased molecular weight. **Top row**: WT- pUC119His6*mycXba* vector, 1-10) $R^{H}50A$ protoplast screens. All +ve. **Bottom row**: WT - pUC119His6*mycXba* vector 1-10) $R^{H}56A$ protoplast screens. All +ve.

5.5.3.2 Expression and purification of the 20G9scFv mutants

Cell lysates from each small-scale expression of the mutants were Western blotted, using the 9E10 antibody (Figures 5.21 & 5.22). Large scale expressions were performed and each scFv IMAC purified (Figure 5.20 & 5.23). However, the $Y^{L}96A$ scFv mutant expressed at a level that was only just detectable using Western blotting. Despite several attempts, insufficient protein was obtained for kinetic characterisation.

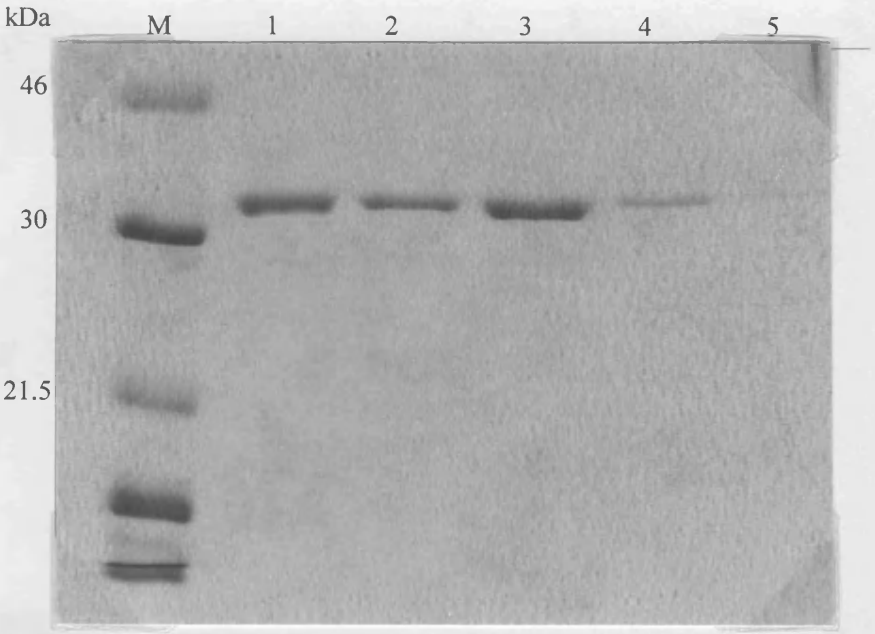


Figure 5.20 12% SDS-PAGE resolution of expressed and purified scFv antibody fragments. 1) 20G9WT scFv (230μg/ml) 2) 20G9-$R^{H50}A$ scFv (220μg/ml) 3) 20G9-$Y^{L91}A$ scFv (235μg/ml) 4) 20G9-$R^{H56}A$ scFv (160μg/ml) 5) 20G9-$Y^{L96}A$ scFv (<20μg/ml). M - Rainbow marker (low M.Wt).

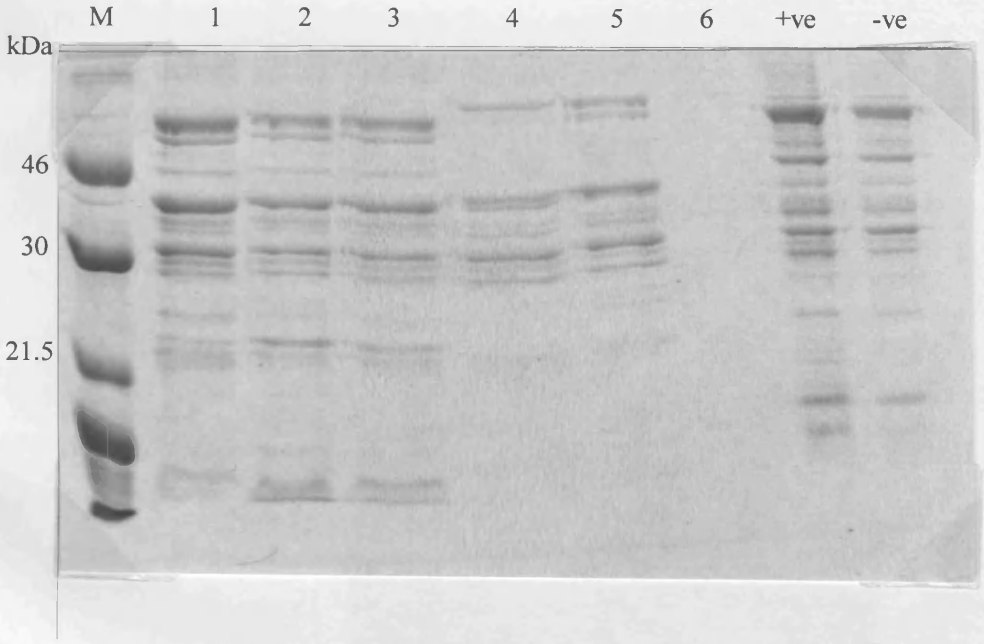


Figure 5.21 Coomassie blue stained 12% SDS-PAGE of 20G9 scFv-WT and mutant cell lysates and eluate. 1) mutant $R^{H50}A$ 2) mutant $Y^{L91}A$ 3) mutant $Y^{L96}A$ 4) 20G9scFv-$WT^{\#1}$ 5) 20G9scFv-$WT^{\#2}$ 6) 20G9scFv-WT eluate. [+ve: *myc*-tagged Fab] [-ve: WT plasmid].

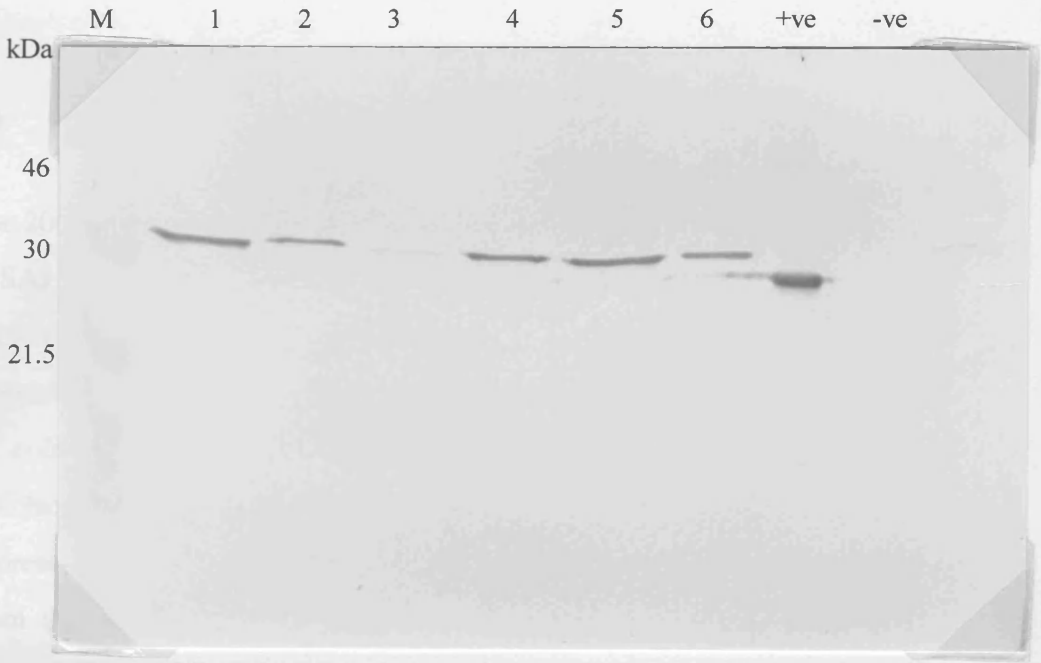


Figure 5.22 Western blot using 9E10 anti-*myc* tag antisera applied to a 12% SDS-PAGE of 20G9 scFv-WT and mutant cell lysates and eluate, identical to the above Coomasie blue stained gel. 1) mutant $R^{H50}A$ 2) mutant $Y^{L91}A$ 3) mutant $Y^{L96}A$ 4) 20G9scFv-$WT^{\#1}$ 5) 20G9scFv-$WT^{\#2}$ 6) 20G9scFv-WT eluate. [+ve: *myc*-tagged Fab] [-ve: WT plasmid].

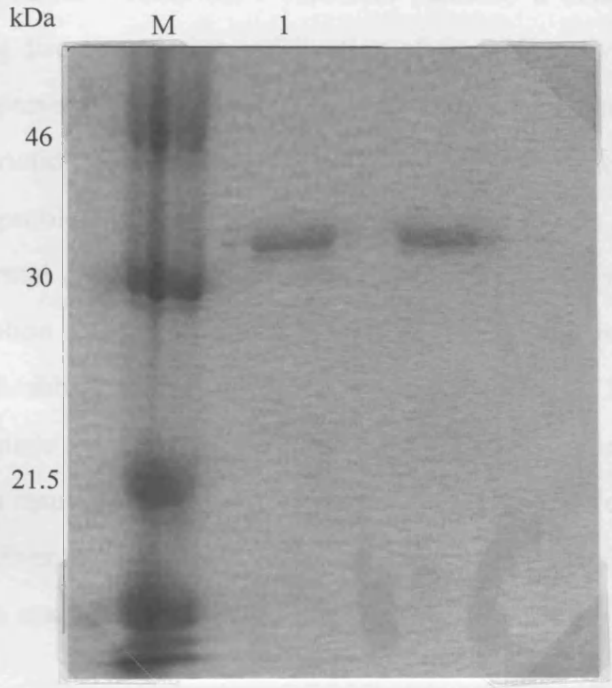


Figure 5.23 12% SDS-PAGE resolution of expressed and purified scFv antibody. 1) 20G9-$Y^{L94}A$ scFv (225μg/ml). M - Rainbow marker (low M.Wt).

5.6 EXPRESSION OF 20G9SCFV: CONCLUSIONS

The 20G9scFv construct, cloned into the Pharmacia vector pKK233-2 by IGEN Inc. (USA) had major problems with expression. Initial expression of the 20G9scFv antibody was in batch culture using the *E. coli* strain XL-1 Blue but was barely detectable on Coomasie blue staining. Other bacterial strains were explored and the *E. coli* strain TG1 was selected since it produced higher levels of 20G9scFv. However, levels of expression in TG1 fell over time to the point where expression ceased. Even cells freshly transformed with construct or cells taken from the initial glycerol stocks failed to reverse the loss of expression. Other reports of instability of this vector were later confirmed by Pharmacia. A decision was taken to clone into a replacement vector donated by Pharmacia, for which the problem of vector instability had been resolved.

The pTrc99A vector was the "improved" successor to pKK233-2 but its use led to other problems. The pKK233/WT20G9scFv construct included a disabled *Nco*I restriction site, preventing direct restriction and ligation of the 20G9scFv gene into pTrc99A. It was first necessary to introduce an *Nco*I restriction site using PCR. However, successful restriction and cloning of the PCR product into the pTrc99A vector introduced a new problem. The *Nco*I restriction site encoded an ATG start codon adjacent to the normal ATG start codon located in the pelB leader sequence of 20G9scFv. Transcription of this construct led to out-of-frame expression of 20G9scFv and required disabling, as seen in the construct, pKK233/TW20G9scFv. The numerous attempts made to delete several of the *Nco*I bases using mung bean nuclease either resulted in insufficient deletion to disable the *Nco*I site or deletion of large amounts of DNA. Eventually the decision was taken to abandon pTrc99A in favour of a vector from an academic source with some history of success.

The vector pUC119His6*mycXba*, donated from Dr. Greg Winter's group, had successfully been used for scFv expression. This vector contained a hexa-histidine tag for purification and a *c-myc* tag for identification. This required removal of the

penta-histidine tag present on the 20G9scFv gene. This was achieved during introduction of the required *Not*I and *Hind*III sites using PCR. However, restriction and ligation of the PCR product into the vector proved problematic with repeated failure to produce constructs. Digestion of the ^{32}P labelled PCR product confirmed that product digestion was occurring. Persistant PCR led to a construct that was verified by sequencing to be free of any mutations. Small scale expression (50ml) was attempted and successful expression of the wild type 20G9scFv was seen using Western blotting towards the *myc* tag. This was followed by scale up and optimisation of expression.

A parallel attempt to clone 20G9scFv into the vector p4xH was performed while cloning into pUC119His6*mycXba*. p4xH is a chimaeric Fab expression vector supplied by Professor Peter Schultz's group (UCSF) and in which strong expression had been observed. The V_H and V_L were both successfully cloned into this vector, but was halted after successful expression in pUC119His6Myc*Xba*.

Expression of the antibody construct in pUC119His6*mycXba* was first optimised and then followed by scale up of expression to produce antibody in quantities sufficient for purification. Optimisation was monitored using the *myc* tag that allows detection of antibody from small-scale expressions. Stronger expression was clearly observed using the *E. coli* strain TOPP2 over TG1. The length of induction was important; strong expression into the periplasmic space was seen at 3 and 24 hours, whereas expression into the supernatant was only seen after 24 hours. No improvements were observed with the use of protease inhibitors. Purification of scFv from the periplasm and supernatant was attempted using Ni^{2+} chelating columns. However, scFv in sufficient quantities was only successfully purified from the periplasm. No scFv was purifiable from the supernatant. This may have been due to the presence of ions within the spent media interfering with the ability of the nickel to chelate the antibody fragments. Subsequent large scale expressions used a 3 hour induction followed by recovery of scFv from the periplasm.

ScFv from the large-scale batches were captured using batch binding. This is where Ni^{2+} chelate resin is added direct to the vessel containing the periplasmic preparation. This method increases the chances of reaching binding equilibrium which is less likely than simply running the scFv through resin packed into a short column. The initial method of elution involved washing in 50mM NaPO4 pH 8.0, 300mM NaCl, 20mM Imidazole, followed by elution of scFv using 0.5M EDTA or 250mM Imidazole added in a single step. However, the scFv produced from this method required further purification. When gel filtration or ion exchange were attempted a large proportion of the antibody was lost, probably due to the relatively low starting levels of antibody in each preparation. Consequently a different approach was required. After the antibody had been batch bound to the Ni^{2+}, the resin was collected into a column and connected to an FPLC system, allowing the use of an imidazole gradient for elution. This allowed far greater control and consequently production of high purity scFv suitable for kinetic studies.

When a reliable method of expression and purification had been established, each mutant construct was cloned into the pUC119His6*mycXba* vector. The levels of expression of each mutant varied quite widely, 20G9WT scFv being the highest, followed by 20G9-R^{H50}A, 20G9-Y^{L91}A, 20G9-R^{H56}A and 20G9-Y^{L94}A. The mutant 20G9-Y^{L96}A failed to express in quantities sufficient for purification, resulting in abandonment of this construct for kinetic characterisation.

CHAPTER 6

6 KINETIC CHARACTERISATION OF FOUR MUTANTS OF 20G9 SINGLE CHAIN FV

6.1 INTRODUCTION

For catalytic antibodies, relatively high levels of antibody are required for kinetic characterisation since their turnover rates are normally much lower than those of enzymes. When using such relatively large amounts of protein it is important to establish that the observed catalysis is due to the antibody and not a contaminant, necessitating purification steps which, in themselves, cause some loss of protein.

The 20G9 antibody was first produced in 1988 [Durfor *et al,* 1988] and catalyses the hydrolysis of the ester, phenyl acetate, resulting in the formation of two products, phenol and acetic acid. Phenyl acetate undergoes hydrolysis in aqueous solution and is thought to pass through a tetrahedral transition state during its breakdown. To generate the murine antibody response, a stable analogue of the proposed transition state thought to approximately mimic its geometry and bond lengths was synthesised (Figure 6.1). The transition state mimic, phenyl phosphonate, was conjugated to keyhole limpet hemocyanin (KLH) and used to immunise BALB/c mice. 20G9 IgG was one of several catalytic antibodies produced from the resulting hybridoma cells. The antibody was purified from the hybridoma ascites fluid using protein A affinity chromatography. This was the material used for the original kinetic characterisation of the wild type antibody [Martin *et al*, 1991a].

Kinetic measurements can be made spectrophotometrically if the extinction coefficient of one of the reactant or product species is known. For 20G9, the phenol produced can be followed at 270nm and its rate of production used to calculate the reaction rate. The substrate, phenyl acetate, spontaneously hydrolyses with a

measurable rate, so it was important to use a dual-beam spectrophotometer to allow subtraction of the background reaction. During catalysis of phenyl acetate by 20G9, an unusual profile is seen where the reaction slows dramatically after a few turnovers. The initial rate is described as the burst rate and following that the steady-state rate. This two-phase profile is possibly due to product inhibition by phenol, supported by the observation that the addition of phenol eliminates the burst [Martin *et al*, 1991a].

As discussed in chapter 1, preliminary kinetic evidence supported the involvement of tyrosine residues within the active site of 20G9 [Martin *et al*, 1991a]. These observations, in part, led to the decision to mutate some of the CDR tyrosine residues. Previous kinetic characterisation of the 20G9 intact IgG determined K_M's of 300 μM and 36 μM and k_{cat}'s of 4.9 min^{-1} and 0.54 min^{-1} for the burst and steady-state rates respectively, [Martin *et al*, 1991a]. The calculated acceleration of reaction over background ($\mathbf{k_{uncat}}$ = 4.9 x 10^{-4} min^{-1}) gave a value of k_{cat}/k_{uncat} = 1.0 x 10^4 min^{-1} for the burst phase and 1.1 x 10^3 min^{-1} for the steady-state phase.
The relatively low rates of reactions catalysed by catalytic antibodies make it important to establish that the observed activity can be attributed to the antibody. For 20G9, this was attempted by the following:

i) observation of the distinctive burst/steady-state reaction profile for 20G9 due to product inhibition by phenol. Contaminating enzymes are likely to have a single linear profile. The burst phase of 20G9 can be eliminated by the addition of phenol.

ii) rates approximate to those previously observed during kinetic characterisation of 20G9.

iii) addition of the transition state analogue to the reaction. The antibody that was raised against the TSA has high affinity towards it, thus the TSA should inhibit catalysis.

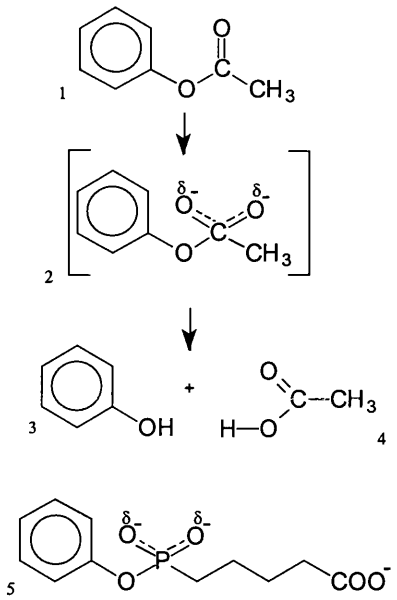


Figure 6.1 The 20G9 antibody catalysed hydrolysis of phenyl acetate [1] producing phenol [3] and acetic acid [4]. The presumed transition state [2] is shown on which the design of the transition state analogue (TSA) [5] was based.

The aim of this chapter to determine if the activities of 20G9 mutant scFvs have been altered by the specific residue changes. The burst and steady-state rates of the whole antibody had been previously characterised, allowing a degree of comparison to be made between the scFv antibody and the whole IgG. The mutant kinetics may allow a model to be proposed in which the function of particular residues within the abzyme's binding site can be attributed a specific role.

6.2 MATERIALS

6.2.1 Preparation of phenyl acetate

Phenyl acetate (Sigma-Aldrich) was stored at 4°C and sealed with parafilm to exclude entry of water vapour. A 50mM solution of phenyl acetate and 250mM dimethyl sulfoxide (DMSO), (Sigma-Aldrich) was first prepared. Phenyl acetate (F.W. 136.15; density 1.073) exists as a liquid at 7.88M. 317.2µl of phenyl acetate was mixed with 900µl of DMSO and made to a total volume of 50ml with MilliQ water. This was shaken by hand initially and then mixed by inversion at 4°C for 1 hour until completely homogenous. 0.5ml of the 50mM solution was taken and diluted to 5ml with MilliQ water with a resulting substrate concentration of 5mM.

The 5mM phenyl acetate solution was stored on ice while in use and discarded at the end of the day.

6.2.2 Antibody preparation

20G9 wild type and mutant scFv antibodies were concentrated in Centricon-10 concentrators (Amicon) and their concentrations measured using a BCA protein assay kit (Pierce). The antibodies were all diluted with catalytic assay buffer to give approximate concentrations of 200-250 μg/ml. The scFv antibodies were dispensed into 60μl aliquots in 0.5ml eppendorfs and snap frozen in dry ice and acetone and stored at -70°C until used. Only freshly thawed aliquots were used for each kinetic study with all antibody aliquots held on ice during the studies.

6.2.3 Dual-beam Spectrophotomer

A 3B dual-beam spectrophotometer (Perkin-Elmer) was used to collect kinetic data from the assays. The instrument was driven by Perkin-Elmer software. Assay data were stored on disk for statistical analysis using Grafit v3.01 (Erithacus Software Ltd.).

6.3 METHODS

6.3.1 Determination of the background hydrolysis (k_{uncat}) of phenyl acetate

Catalytic assay buffer and two 0.5ml optically matched quartz cuvettes were allowed to equilibrate at 25°C in a water bath. Up to 500μl of catalytic assay buffer (10mM Tris-HCl pH 8.8, 140mM NaCl) was added to the reference and sample cuvettes. Readings were taken over a range of phenyl acetate concentrations 0-5000 μM. The 5mM phenyl acetate solution was throughly mixed before addition to the cuvette to ensure complete homogeneity. Absorbance at A_{270} was measured at 5 second intervals over a period of 25 minutes. Zero substrate measurements confirmed the correct functioning of the dual beam spectrophotometer and eliminated drift. Regular checking of the substrate k_{uncat} ensured that the prepared substrate was consistent between assays.

6.3.2 Catalytic assays on 20G9 wild-type scFv and its mutants

The catalytic assay buffer and the 0.5ml cuvettes were allowed to warm to 25°C in a water bath. Appropriate volumes of pre-warmed assay buffer were added to the reference and sample cuvettes. Each scFv antibody was added to the sample cuvette, inverted and allowed 30 seconds to equilibrate in the buffer. Phenyl acetate was then added to the cuvettes, mixed by inversion and read at A_{270} over a period of up to 25 minutes at 5 second intervals. Assays were performed over a range of phenyl acetate concentrations (100, 200, 300, 400, 500, 600 μM). An assay was repeated into which 12.5μM phenol (Sigma-Aldrich) was added the the sample and reference cuvettes to allow observation of any changes to the reaction profile through product inhibition.

6.4 RESULTS

6.4.1 Background hydrolysis of phenyl acetate

No increase in absorbance was seen when both cuvettes contained only assay buffer or when equal quantities of phenyl acetate were added to both cuvettes. This verified correct functioning of the dual-beam spectrophotometer. When the concentration of phenyl acetate was increased the absorbance increased in direct proportion (Table 6.1). V ($\mu Mmin^{-1}$), the rate of hydrolysis, was calculated using an extinction coefficient of $A\varepsilon_{270} = 1425\ M^{-1}\ cm^{-1}$ for phenyl acetate. Linear regression was performed on a plot of V ($\mu Mmin^{-1}$) against [S] (μM) (Figure 6.2). The gradient of the line ($y = ax + b$) gave a gradient of $5.6 \times 10^{-4}\ min^{-1}$ which equates to the k_{uncat}.

Table 6.1 First order hydrolysis of phenyl acetate. Increasing substrate concentration (μM) resulted in a proportional increase in rate (V) of phenol production. The slope of the linear regression is taken as the k_{uncat}.

Phenyl acetate (μM)	V (μM min^{-1})
0	0.0
100	0.212
200	0.297
400	0.410
800	0.552
1000	0.764
1200	0.850
1600	1.146
2000	1.570
2400	1.485
2800	1.824
3200	1.740

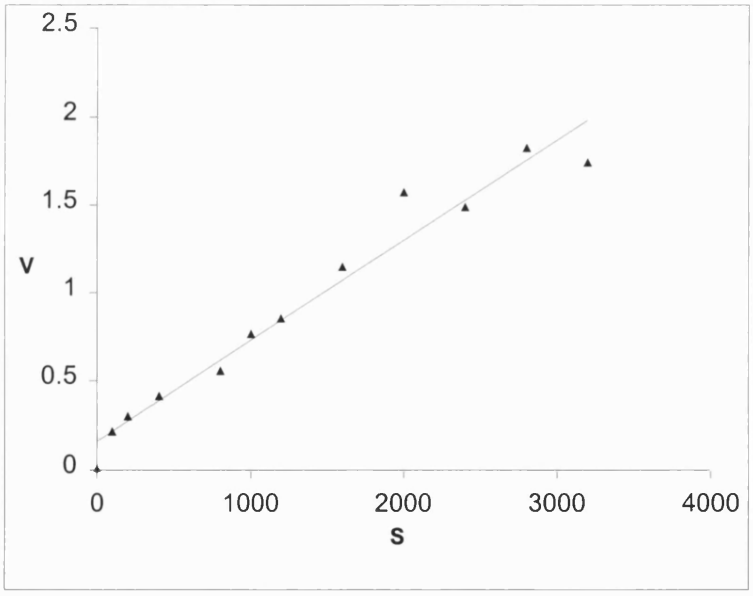
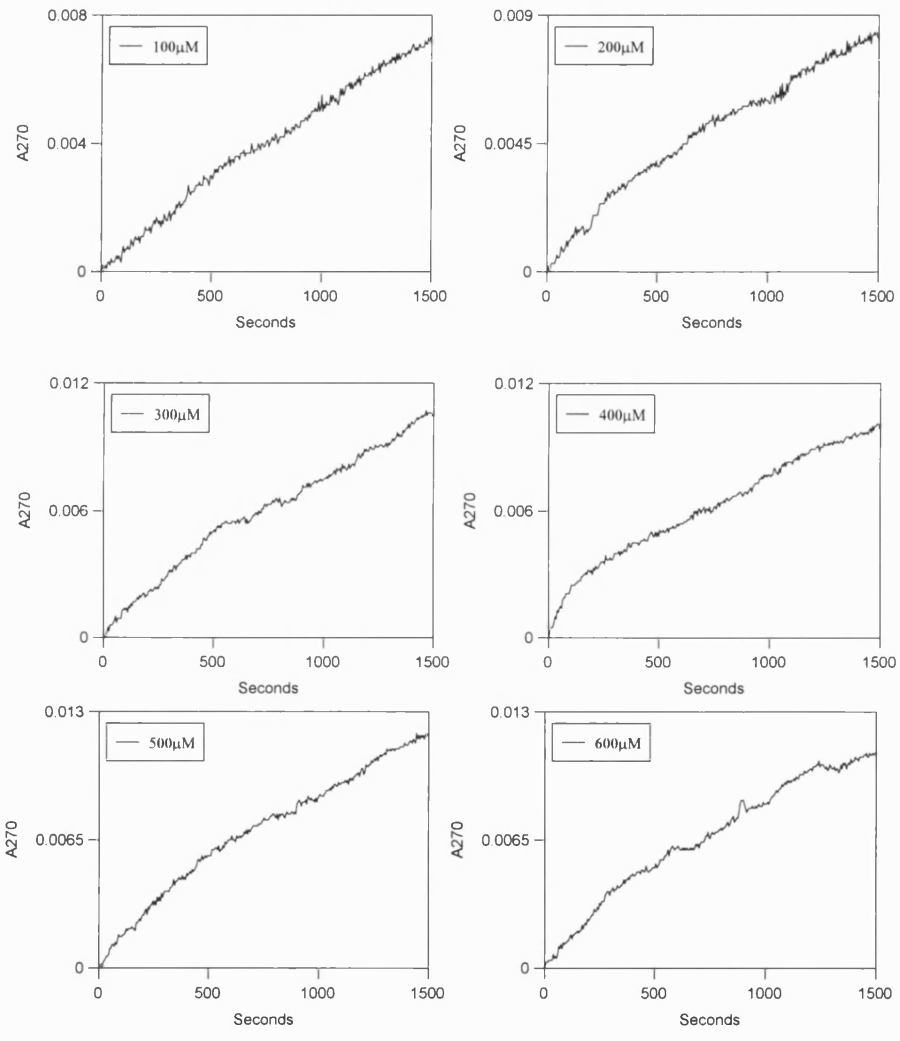


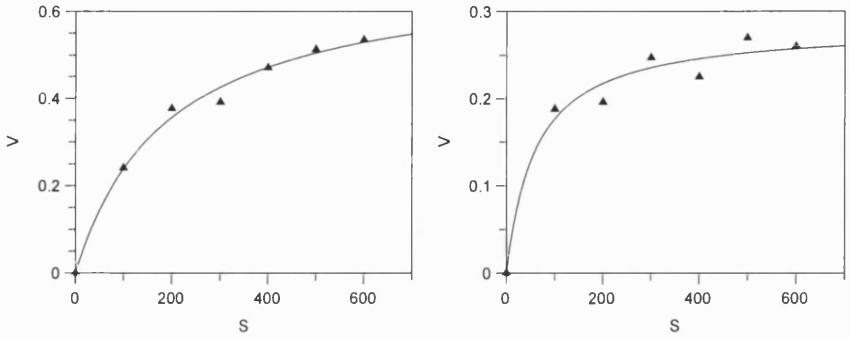
Figure 6.2 V (μM min^{-1}) vs [S] (μM) from the spontaneous hydrolysis of phenyl acetate. Linear regression performed on the data gave a gradient of 5.6×10^{-4} min^{-1}.

6.4.2 Determination of the K_M, V_{max} and k_{cat} of 20G9 wild type scFv

Six assays were performed on 20G9 wild-type scFv (0.51μM per assay) and all displayed a visible burst and steady-state phase. The gradients (A_{270} sec^{-1}) for each of the experiments were calculated for the burst phase (0-320 seconds) and steady-state phase (320-1495 seconds) by linear regression for each substrate concentration. The gradient values (A_{270} s^{-1}) were divided by the extinction coefficient (A_ε) for phenyl acetate at pH8.8 (1425 M^{-1} cm^{-1}) to convert the slope to a rate value (V μM min^{-1}) (Table 6.2).



The rate (V) was plotted against the phenyl acetate concentration [S] used in each assay (Table 6.2) and a simplex curve fitted to the data from which the K_M and V_{max} were calculated to 95% confidence intervals.



Burst phase Steady-state

The k_{cat} for the catalytic antibody was calculated using the V_{max} and the antibody concentration using the formula $k_{cat} = V_{max}/[Ab_0]$.

Table 6.2 Linear regression was applied to the 20G9 wild-type scFv data points of the burst and steady-state phases from which the gradient (A_{270} sec^{-1}) was calculated. The extinction coefficient of phenyl acetate [$A\varepsilon_{270}$ = 1425 M^{-1} cm^{-1}] was used to convert the gradient to V (μM min^{-1}) [V=slope/A_ε].

	Burst	Steady-state
[S] (μM)	V (μM min^{-1})	V (μM min^{-1})
0	0.000	0.000
100	0.241	0.188
200	0.377	0.197
300	0.392	0.247
400	0.471	0.238
500	0.513	0.270
600	0.535	0.260

A burst and post-burst phase was seen on all the assays performed. The characteristic burst and steady-state profile seen with the hybridoma derived 20G9 and its elimination by the addition of 12.5μM phenol to the assay was reproduced with 20G9 scFv (Figure 6.3), and suggested that the hybridoma derived 20G9 and the 20G9 scFv were behaving in a comparable manner.

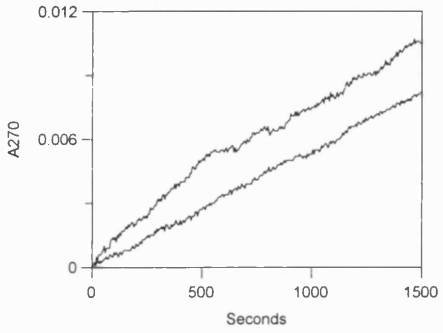


Figure 6.3 Kinetic profile of 20G9scFv at a phenyl acetate concentration of 300uM. The lower profile is from an assay to which 12.5uM phenol has been added.

The values obtained for the Michaelis constant, K_M and the catalytic constant, k_{cat}, for the 20G9 scFv wild-type during the burst phase were 193±31 μM, and 1.37±0.07 min^{-1} and are comparable to the values previously obtained for the hybridoma derived 20G9 antibody (K_M = 300 μM and k_{cat} = 4.9). The K_M and k_{cat} values obtained for 20G9 scFv wild-type during the steady-state phase were 63±18 μM and 0.56±0.08 min^{-1} respectively, again similar to the values obtained for the hybridoma derived 20G9 antibody (K_M=24.4 μM and k_{cax}=0.54 min^{-1}) (Table 6.3). The acceleration (k_{cat}/k_{uncat}) of the hydrolytic reaction by 20G9 wild-type scFv over background (k_{uncat} = 5.9 x 10^{-4} min^{-1}) is 2.3 x 10^3 for the burst and 9.5 x 10^2 for the steady-state.

Table 6.3 The catalytic and binding properties of the 20G9 scFv compared to those of the hybridoma derived 20G9 antibody.

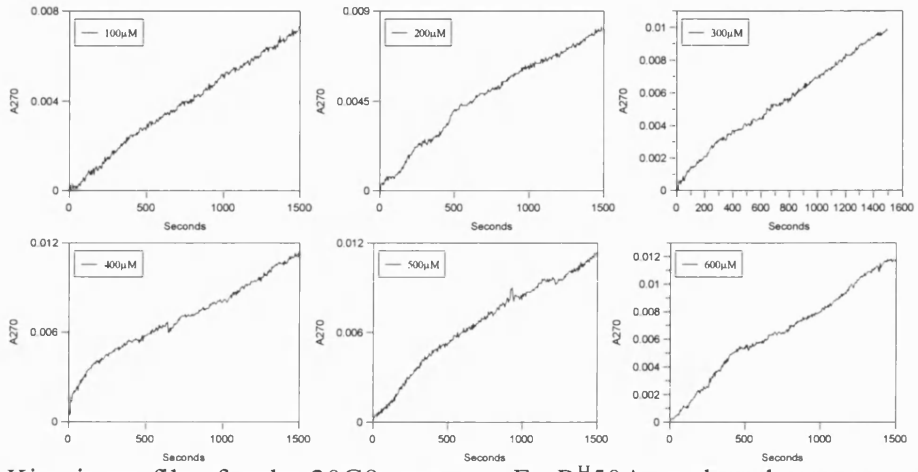
	Phase	V_{max} ($\mu M\ min^{-1}$)	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
20G9 scFv	Burst	0.699±0.040	1.37±0.07	193±31	7.1×10^3
	Steady-state	0.288±0.04	0.56 ±0.08	63±18	8.9×10^3
20G9 IgG	Burst	-	4.9	300	1.6×10^4
	Steady-state	-	0.54	24.4	2.2×10^4

When the data of the scFv and hybridoma derived antibody are compared, the catalytic and binding properties are relatively similar. However, the hybridoma derived antibody clearly has a higher catalytic turnover during the burst phase compared to that of the scFv, while the K_M of the scFv is lower. During the steady-state phase, the reverse is true with the scFv displaying slightly weaker binding to the substrate but having a marginally higher catalytic turnover when compared to the hybridoma derived antibody. The differences may be due to the greater stability of the whole IgG when compared to that of a scFv. The IgG may provide a superior environment for the transition state of the hydrolytic reaction through the structural stability of the Fv region, but in turn may have a reduced affinity towards the substrate or ground state. In contrast, scFv may have a higher affinity for the substrate or ground state due to its greater flexibility, possibly due to the absence of the constant regions. The disadvantage of any additional flexibility in the scFv is that it may reduce the effectiveness of catalysis. During the steady-state phase, the differences between the catalytic and binding properties of the hybridoma derived antibody and the scFv are negligible. The product inhibition by phenol may act by reducing the effects that any structural differences have upon binding and catalysis through steric hinderance within the active site of the 20G9 IgG and 20G9 scFv.

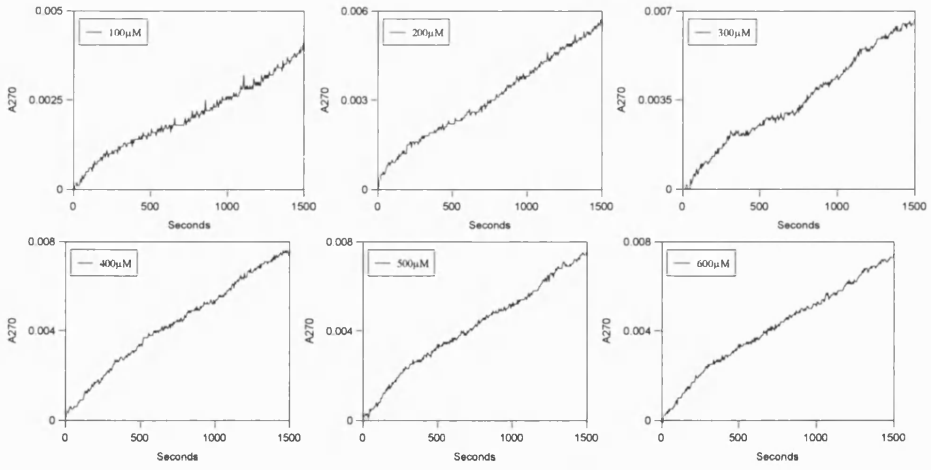
6.4.3 Determination of the K_M, V_{max} and k_{cat} of the 20G9 scFv mutants, R^H50A, R^H56A, Y^L91A, Y^L94A

Six assays were performed upon each of the 20G9 scFv mutants R^H50A, R^H56A, Y^L91A and Y^L94A at scFv concentrations as close a possible to those used in the 20G9 scFv wild-type catalytic assays (0.48, 0.35, 0.51 and 0.49μM respectively) to

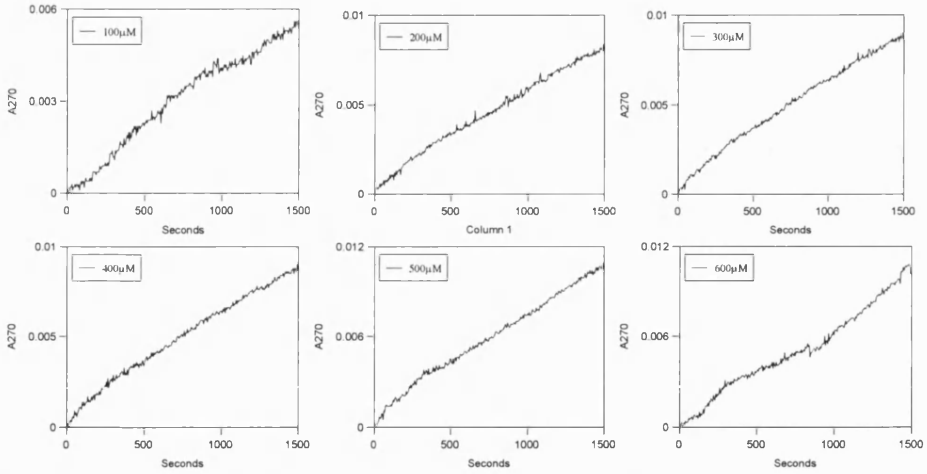
ensure comparability of the assays thus allowing any differences to be attributed to their binding and catalytic properties. Each mutant displayed a burst and steady-state phase. As for the wild-type scFv, the slope for each assay (A_{270} sec^{-1}) was taken for the burst phase over 0-320 seconds and for the steady-state phase over 320-1495 seconds. The slope was calculated for each phase using linear regression and the gradient values (A_{270} s^{-1}) divided by the phenyl acetate extinction coefficient (A_{ε}) (1425 M^{-1} cm^{-1}) converting the slope to V μM min^{-1}.



Kinetic profiles for the 20G9 mutant scFv $R^{H}50A$ at phenyl acetate concentrations of 100μM, 200μM, 300μM, 400μM, 500μM & 600μM.

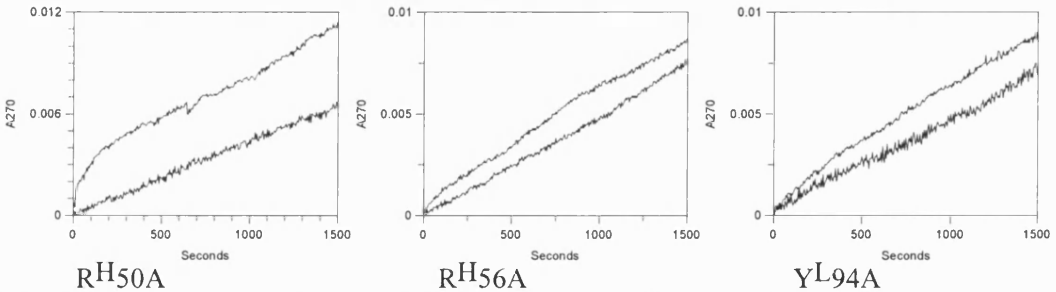


Kinetic profiles for the 20G9 mutant scFv $R^{H}56A$ at phenyl acetate concentrations of 100μM, 200μM, 300μM, 400μM, 500μM & 600μM.



Kinetic profiles for the 20G9 mutant scFv $Y^{L}94A$ at phenyl acetate concentrations of 100µM, 200µM, 300µM, 400µM, 500µM & 600µM.

Elimination of the characteristic burst and steady-state profile through the addition of 12.5µM phenol was repeated with each mutant.



$R^{H}50A$ $R^{H}56A$ $Y^{L}94A$

Figure 6.4 Kinetic profile of the 20G9 scFv mutant $R^{H}50A$, $R^{H}56A$, $Y^{L}94A$ at a phenyl acetate concentration of 400µM with and without the addition of 12.5µM Phenol to the reaction.

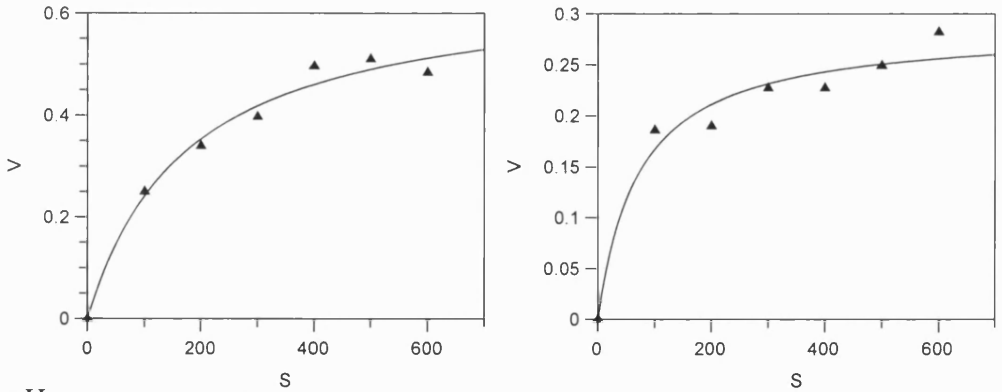
Each mutant saw the burst phase of the profile eliminated on the addition of 12.5µM phenol to the assay. However, it was observed that the characteristic burst/steady-state profile observed for 20G9 wild-type and for the 20G9 $R^{H}50A$ scFv mutant were less pronounced in $R^{H}56A$ and $Y^{L}94A$ mutant scFvs (Figure 6.4).

Table 6.4 Linear regression was applied to the six kinetic profiles for each mutant scFv, $R^{H}50A$, $R^{H}56A$, $Y^{L}91A$, $Y^{L}94A$, which allowed the gradient for the burst and steady-state phases to be calculated based upon the A_{270} sec^{-1}. The extinction coefficient of phenyl acetate [$A\varepsilon_{270}$ = 1425 M^{-1} cm^{-1}] was used to convert the gradient (A_{270} sec^{-1}) to a rate, **V** (μM min^{-1}), [V=slope/A_{ε}], for each of the substrate concentrations used. ND - No activity observed.

	$R^{H}50A$		$R^{H}56A$		$Y^{L}91A$		$Y^{L}94A$	
[S] μM	Burst V μM min^{-1}	Steady-state V μM min^{-1}	Burst V μM min^{-1}	Steady-state V μM min^{-1}	Burst V μM min^{-1}	Steady-state V μM min^{-1}	Burst V μM min^{-1}	Steady-state V μM min^{-1}
0	0.000	0.000	0.000	0.000	NA	NA	0.000	0.000
100	0.250	0.186	0.158	0.095	NA	NA	0.173	0.138
200	0.340	0.190	0.203	0.146	NA	NA	0.297	0.210
300	0.397	0.227	0.295	0.180	NA	NA	0.326	0.224
400	0.496	0.227	0.291	0.184	NA	NA	0.350	0.222
500	0.510	0.249	0.335	0.182	NA	NA	0.425	0.271
600	0.484	0.282	0.340	0.176	NA	NA	0.401	0.275

The rate (V) was plotted against the phenyl acetate concentration (S) used in each assay (Table 6.4) and a simplex curve fitted to the data from which the K_M and V_{max} were calculated to 95% confidence intervals. The k_{cat} for the catalytic antibody was calculated using the V_{max} and the antibody concentration using the formula $k_{cat} = V_{max}/[Ab_0]$.

6.4.3.1 $R^{H}50A$ scFv

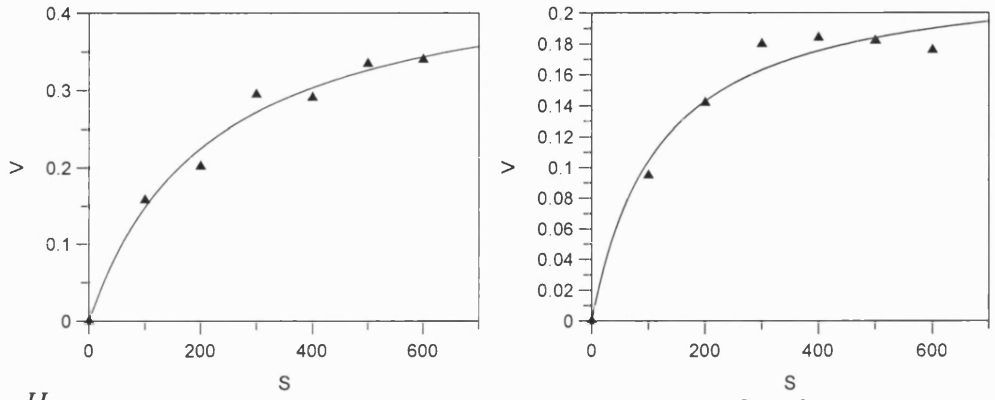


$R^{H}50A$ scFv Burst *Steady-state*

The values obtained by the Simplex curve fit for the the Michaelis constant, K_M and the catalytic constant, k_{cat} of 20G9 $R^{H}50A$ scFv mutant were 173±40 μM and k_{cax}=1.25±0.11 min^{-1}. The K_M and k_{cat} values obtained for the steady-state phase were 71±27 μM and 0.59±0.04 min^{-1} respectively. The acceleration (k_{cat}/k_{uncat}) of the

hydrolytic reaction by the mutant scFv is by a factor of = 2.1 x 10^3 for the burst and 1.0 x 10^3 for the steady-state.

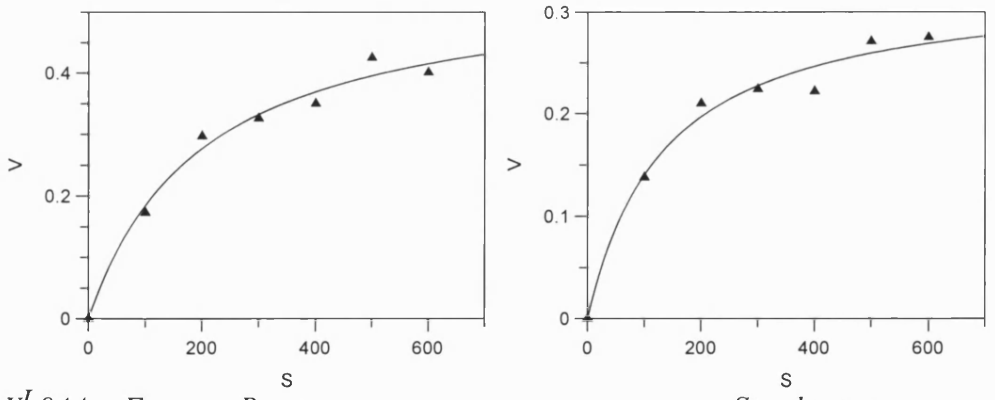
6.4.3.2 R^H56A scFv



R^H56A scFv Burst *Steady-state*

For the burst phase, the K_M and k_{cat} of 20G9 R^H56A scFv mutant was 213±48 μM and k_{cax}=1.33±0.11 min^{-1} respectively. The K_M and k_{cat} values obtained for the steady-state phase were 113±32 μM and 0.64±0.04 min^{-1} respectively, an acceleration over the background reaction of 2.25 x 10^3 for the burst and 1.1 x 10^3 for the steady-state.

6.4.3.3 Y^L94A scFv



Y^L94A scFv Burst *Steady-state*

The K_M and k_{cat} of the Y^L94A mutant for the burst phase was 199±45 μM and k_{cax}=1.12±0.08 min^{-1} respectively. The K_M and k_{cat} values for the steady-state phase were 134±32 μM and 0.67±0.05 min^{-1} respectively, an acceleration over the background reaction of 1.9 x 10^3 for the burst and 1.1 x 10^3 for the steady-state.

6.4.3.4 $Y^{L}91A$ scFv

A number of assays were performed on the mutant $Y^{L}91A$ antibody upon several occasions but no measurable activity was detected. The mutant expressed well (Figures 5.20 and 5.22) and were easily picked up by Western blotting. During the assays, the concentration of scFv added to the assay was increased up to 0.9μM in an attempt to demonstrate activity, yet no activity was observed.

Table 6.5 Comparison of the catalytic and binding properties of the 20G9 scFv mutants and wild-type. NA indicates that no activity was observed.

scFv	Phase	V_{max} ($\mu M\ min^{-1}$)	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
20G9-W	Burst	0.699±0.040	1.37±0.07	193±31	7.1×10^3
	Steady-state	0.288±0.04	0.56 ±0.08	63±18	8.9×10^3
$R^{H}50A$	Burst	0.600±0.051	1.25±0.11	173±40	7.2×10^3
	Steady-state	0.286±0.022	0.59±0.04	71±27	8.3×10^3
$R^{H}56A$	Burst	0.466±0.040	1.33±0.12	213±48	6.2×10^3
	Steady-state	0.226±0.017	0.64±0.04	113±32	5.7×10^3
$Y^{L}94A$	Burst	0.553±0.046	1.12±0.08	199±45	5.6×10^3
	Steady-state	0.329±0.023	0.67±0.05	134±32	5.0×10^3
$Y^{L}91A$	Burst	NA	NA	NA	-
	Steady-state	NA	NA	NA	-

6.4.4 Mutated residues and their effect upon activity

A model of the key residues within the 20G9 active site can be built based upon the: i) activity seen with each mutant ii) the 20G9 Fv *AbM* model iii) kinetic studies upon the 20G9 hybridoma derived antibody iv) residues used in other esterolytic antibodies. The change in activity seen by mutating the arginine $^{H}50$ residue to alanine is negligible (Table 6.5) and so it is extremely unlikely to play a significant role in the catalysis of the phenyl acetate substrate despite its proximity to the putative 20G9 active site. Arginine residues have been shown to form hydrogen bonds with substrates of other catalytic antibodies, such as in 48G7 [Wedemayer *et*

al, 1997], but there is no kinetic evidence to support its involvement in the active site of 20G9.

Table 6.6 The percentage increase/decrease in K_M & k_{cat} for each mutant with respect to the 20G9WT scFv.

	Steady-state		Burst	
	K_M	k_{cat}	K_M	k_{cat}
R^H50A	+13%	+5%	-10%	-9%
R^H56A	+80%	+14%	+10%	-3%
Y^L94A	+112%	+2-%	+3%	-18%
Y^L91A	Not Determined	-100%	Not Determined	-100%

However mutation of the the arginine residue at $^H56$ to alanine resulted in an 80% increase in the value of K_M during the steady-state phase, indicating the affinity of the substrate for the mutant had been reduced, although no significant change in catalytic turnover was seen (Table 6.6). The burst K_M and k_{cat} were largely unaffected. Based upon this evidence it is possible that the arginine residue at $^H56$ is involved in the binding of the substrate, phenyl acetate. It was observed from the 20G9 model that the arginine residue is within the active site groove of the antibody, although the orientation of the side chain in the model oriented away from the active site (Figure 6.5). However, if the orientation is changed, the side chain could face into the groove stabilising the oxygen atoms of the substrate by H-bonding. A similar stabilisation has been seen in the 48G7 catalytic antibody. This stabilising affect might become redundant after the formation of the reaction product, phenol, which itself appears to assist in the binding of the substrate to 20G9 during the steady-state phase.

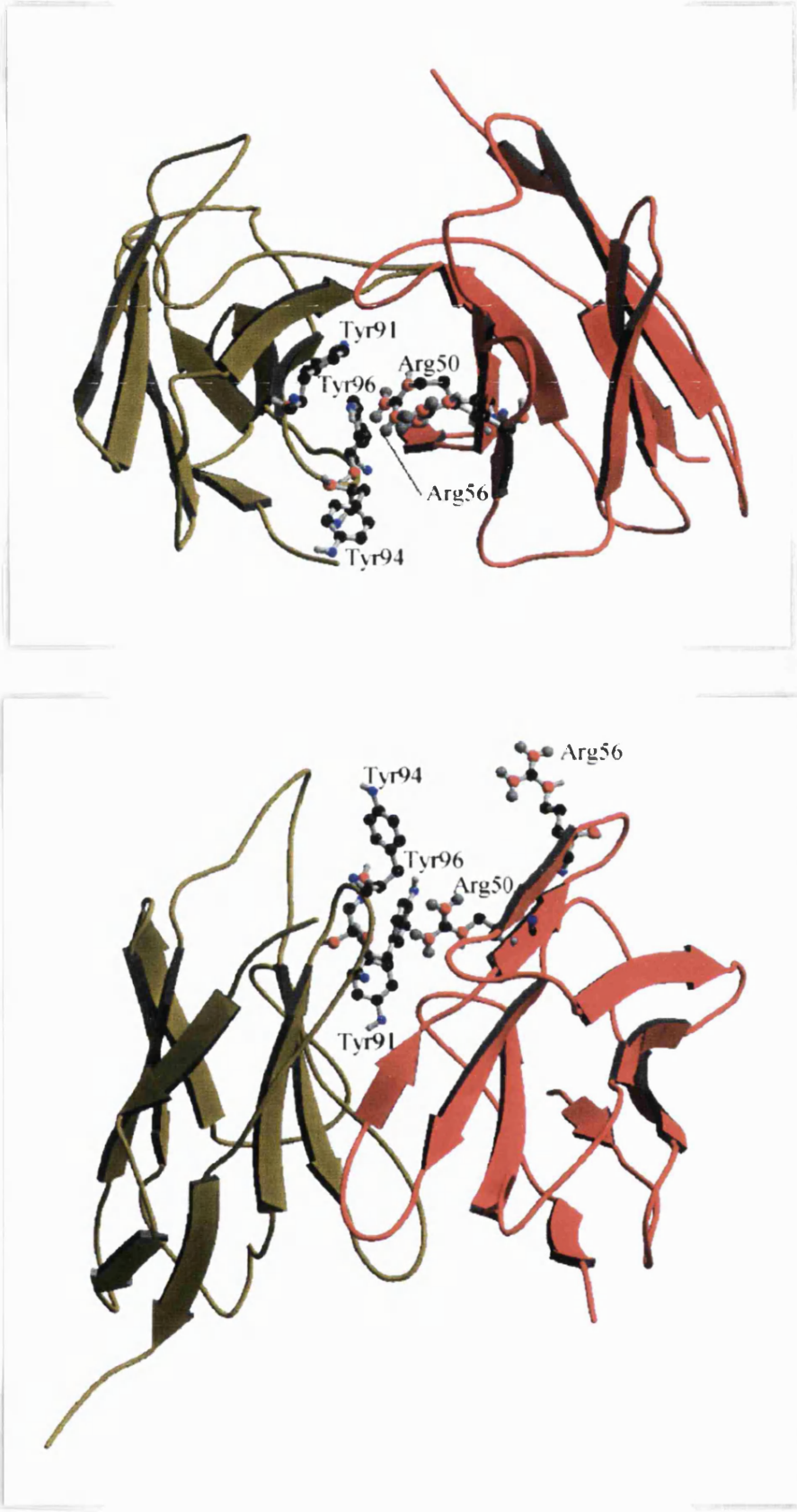


Figure 6.5 Side and topview of the 20G9 Fv model of 20G9Fv displaying the five residues that were mutated ($R^{H50}A$, $R^{H56}A$, $Y^{L91}A$, $Y^{L94}A$, $Y^{L96}A$ & $Y^{L91}A$).

A similar affect is seen in the $Y^{L}94A$ mutant with an increase of 112% in the steady-state K_M value, again indicating a significant reduction in the affinity of the mutant antibody fragment towards phenyl acetate (Table 6.6). As with the $R^{H}56A$ mutant, the steady-state phase k_{cat} and the burst K_M and k_{cat} are comparable to the those seen in the 20G9 WT scFv during the burst and steady-state phases. This would indicate that the tyrosine residue is involved in binding to the substrate. This may be through π-stacking of its aromatic ring with the phenol group of phenyl acetate. However, after several catalytic turnovers, multiple phenols, formed as reaction products, may become involved in π-stacking with the tyrosine side chain(s), hence enhancing binding of the substrate to the active site. This would account for the increased K_M value seen between $Y^{L}94A$ and the 20G9 WT. The comparable k_{cat} values seen during the burst phase and the steady-state phase would indicate that tyrosine at $^{L}94$ has no involvement in the transition state. Interestingly, Tyr^{94L} residue is also found in the active site of the catalytic antibody 48G7, oriented towards the carbonyl atom of the substrate. Thus, Arg^{H56} and Tyr^{94L} may both play an important role in binding phenyl acetate within the antibody's active site groove.

The most dramatic change in kinetic activity was observed with the $Y^{L91}A$ mutant scFv. No catalytic activity was observed with this mutant indicating that this residue has a critical role in the active site of the 20G9 catalytic antibody. This is supported by the model in which the which the Tyr^{91L} residue is seen deep in the active site groove of 20G9, although the model orientation may be incorrect. However, the previous kinetic studies on the hybridoma derived 20G9 IgG showed that a tyrosine was pivotal through the formation of an acetyltyrosyl intermediate [Martin *et al*, 1991a]. This is further reinforced by the crystal structure of 48G7 that shows contact of this same residue with the nitrophenyl ring of the substrate [Patten *et al*, 1996]. These arguments are consistent for the Tyr^{91L} role in the catalysis of phenyl acetate by 20G9.

From the foregoing, a possible mechanism for 20G9 is that the phenol ring of phenyl acetate is buried in the binding site and is stabilised by the Tyr^{94L} sidechain. Other surrounding tyrosine residues (e.g. Tyr^{96L}) may also be involved. Arg^{56H} may H-bond to oxygen atoms on the substrate and is correctly oriented to do so, unlike Arg^{50H}. Hydrolysis at the exposed carbonyl carbon (δ+) occurs possibly due to attack by water or by the Tyr^{91L} phenolate releasing a phenol molecule. Where an acyl intermediate with Tyr^{91L} is formed, this is followed by further hydroxy attack by water resulting in the formation of acetic acid. However, the phenol product does not immediately disperse, possibly due to π-stacking of the phenol with the aromatic rings of the surrounding tyrosines. This acts to stabilise subsequent entry of phenyl acetate molecules into the active site but slows catalysis by steric hindrance or possibly due to slight orientation changes in the positioning of the substrate (Figure 6.6).

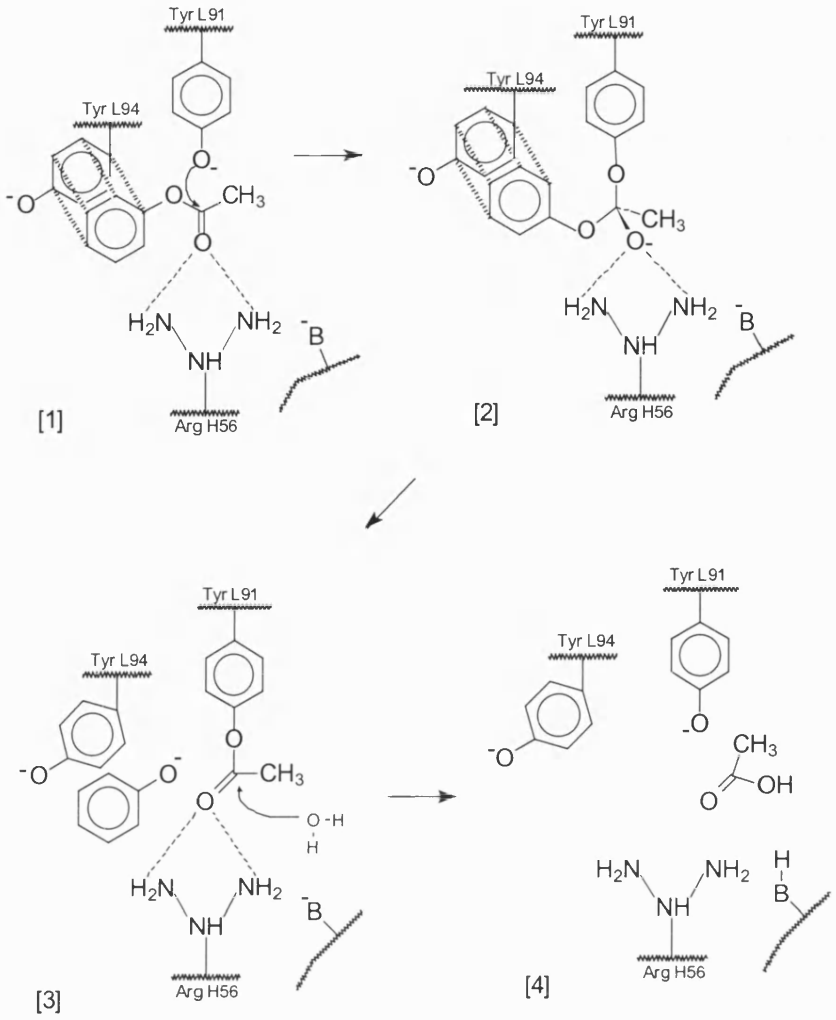


Figure 6.6 Proposed mechanism of the catalytic hydrolysis of phenyl acetate by 20G9.

CHAPTER 7

7 DISCUSSION

The aim of this thesis was to attempt to kinetically characterise the catalytic antibody 20G9, partly based upon the previous biochemical and kinetic characterisation but by producing mutants guided by a model of the antibody. This allowed an attempt to be made to identify those residues in the active site of 20G9 that may have had a role in the antibody's esterase activity.

The 20G9 Fv model

AbM, the antibody modelling package, produced a three-dimensional structure of the antibody Fv and its binding site. All loops excluding H3 were canonical. H3 consisting of only five residues, was modelled by an *ab initio* method. The small size of H3 gave reason to believe that the final model was an accurate representation of the 20G9 Fv. In addition to the model, previous characterisation had implicated tyrosine residues in the catalytic mechanism of the antibody. This, along with evidence of tyrosine and arginine involvement in other esterolytic antibodies, led to a particular focus on these residue types located within the distinctive groove seen in the model of 20G9.

The orientation of the residues within the active site caused the most concern. A decision had to be taken based on whether the orientation of the residues within the model were accurate. The modelling of side chains is possibly the most difficult task for a modelling procedure to perform correctly. Therefore it was decided that provided the Cα atom positions of the candidate residue types were within a cut-off distance of the groove (≈ 15Å), then the side chain would be included.

For practical purposes, residues selected for mutation were split into two groups with those thought most likely to be involved mutated first. The first group

consisted of three heavy chain arginine residues and three light chain tyrosine residues. The second included four light chain tyrosine residues, a heavy chain histidine residue and a heavy chain aspartate residue.

Mutagenesis studies

For the mutation of the selected residues, there was a choice between using a kit based on single oligonucleotide mutagenesis or simply using PCR site-directed mutagenesis. The latter method has previously tended to be overlooked due to the relatively high probability of introducing base mutations into the DNA sequence through the polymerase. However, the arrival of high fidelity DNA polymerases, such as Vent DNA polymerase, has reduced this risk due to their low error rates. One major advantage of PCR mutagenesis over single oligonucleotide mutagenesis is that the degree of control over the DNA is much higher. The use of bacteria in producing the mutant DNA is also eliminated, simplifying mutagenesis.

When working with PCR, it is crucial to optimise the design of the priming oligonucleotides. When designed poorly, they may anneal non-specifically to template DNA, to other oligonucleotides within the reaction, or to themselves. This can have the effect of producing multiple bands of PCR product, smearing, or reduction of reaction efficiency due to the loss of available oligonucleotides within the reaction, sometimes to the point where no product can be seen. In this study, oligonucleotides were computer checked for binding towards the DNA vectors used, the 20G9 gene, and for dimerisation towards partner oligonucleotides or themselves. As a result, almost all PCR reactions performed were successful in producing strong and clean product bands. However, this design process was not foolproof. During the design of the priming oligonucleotides for the mutation of $R^{H}58$, the first forward oligonucleotide design for this mutant (MR58A3) was predicted to dimerise with itself and was subsequently redesigned. However, this oligonucleotide, although apparently satisfactory in design, repeatedly gave an unexpected PCR product size and so this mutant was eventually abandoned.

All of the final PCR products containing the mutated bases within the 20G9 gene were fully sequenced to verify that no polymerase related errors had been introduced before the mutated gene was ligated into the DNA vector. No errors were seen, validating this method as a robust, simple and effective way of introducing specific mutations into a gene.

Expression of 20G9 scFv

The problems encountered with expression of 20G9 scFv are not uncommon in this field. It was very unfortunate that the instability of the vector pKK233-2 could not have been predicted, since a good deal of time was wasted. The vector, pUC119His6*mycXba*, from a non-commercial source was available and had been fully characterised through its extensive use in scFv expression. Obtaining this vector allowed the project to progress after several failures such as with the pTrc99A vector.

Even when using a reliable vector, expression of recombinant antibody molecules and antibody fragments can be difficult. Currently it is impossible to predict whether an antibody will easily express in *E. coli,* and this is especially true for antibody fragments. ScFv's may be particularly difficult to express when compared to whole antibodies or Fab fragments. The groups of Thomas Scanlan and Peter Schultz saw greatly improved expression of their antibody fragments when expressing them as chimeric Fab fragments through the p4xH vector instead of as scFvs. The reason for this is unclear, but may be due to the influence of the constant regions on folding of the antibody fragments molecules during expression, reducing aggregation due to poor folding or misfolded molecules.

When ligating the PCR inserts into the pUC119His6*mycXba* vector, a novel method was used for screening vectors for a PCR insert. Protoplast screening was used based on the differential migration of supercoiled plasmids which had taken up an inserted fragment. This was a rapid method of determining whether ligation had been successful by taking plaques from an agar plate, and digesting the bacterial cell wall resulting in a bacterial protoplast. The addition of a lysis buffer

osmotically shocked the cell, spilling the cytoplasm containing the plasmid. The plasmid, when compared to a plasmid containing no insert, revealed whether successful ligation had been achieved with the whole was used to establish the successful ligation of 20G9 Fv and its 5 mutants into the pUC119His6*mycXba* vector.

20G9 scFv and its mutants were successfully expressed using the pUC119His6*mycXba* vector, although their levels varied even though only a single residue change was present on each mutant. This resulted in the failure to express any of the $Y^{L96}A$ scFv mutant. The presence of the *myc* tag on the antibody fragments allowed small scale expression to be easily followed using Western blotting, while the hexa-histidine tag allowed relatively simple and effective purification of the scFvs.

20G9 scFv kinetic studies

Limited kinetic data were produced from each mutant although they were sufficient to allow calculation of k_{cat} and K_M. An approximate antibody concentration of 0.5μM was used in each kinetic reaction although it was not possible to determine the exact amount of active or correctly folded scFv added to each assay by active-site titration due to the degradation of the transition state analogue, phenyl phosphonate. Proton NMR and phosphorus NMR (400Mhz) analysis, fast atom bombardment mass spectrometry and infrared spectrometry performed on the transition state analogue in the Department of Chemistry at the University of Bath showed that the transition state analogue had decomposed. This was possibly due to hydrolysis by water vapour over its long period of storage (synthesised in 1988 at IGEN Inc. USA). The kinetic behaviour of the wild type scFv was essentially the same as for 20G9. The elimination of the burst phase in the scFv kinetic profile, seen in the hybridoma derived 20G9 antibody (Figure 6.4) was also good evidence of the integrity of the active site of the scFv. The kinetic data for the mutants has allowed a tentative reaction mechanism to be proposed.

During the kinetic studies, differences the background rate of phenyl acetate hydrolysis were observed with a calculated k_{uncat} of 5.6×10^{-4} min^{-1}. The k_{uncat}

determined during the characterisation of 20G9 IgG had a rate of 4.9 x 10^{-4} min^{-1} [Martin *et al*, 1991; Blackburn *et al*, 1991]. It is not clear why this slight difference in rate was observed, but is possibly due to experimental error during measurement.

The 20G9 catalytic mechanism

Stabilisation of the substrate phenyl ring by the aromatic groups (Tyr^{L94} and possibly Tyr^{L96}) is an attractive suggestion for the 20G9 mechanism. Arg^{H56} is in a position where is can stabilise the oxygen(s) of the substrate molecule, as seen in other esterolytic antibodies. Whether the orientation of the Tyr^{91L} phenolate is correct for interaction with the δ^+ charge upon the phenyl acetate carbonyl carbon is unknown but is also possible. This residue when mutated led to elimination of all activity. Previously obtained kinetic data suggested the formation of a tyrosylacetyl intermediate. This has been seen in antibody 48G7 with the same tyrosine (Y^{L91}), lending further support to this hypothesis. These arguments provide strong circumstantial evidence that L91 tyrosine forms a catalytic acyl intermediate.

Future work upon 20G9

Future studies could see completion of the single residue mutants proposed in this thesis. However, only X-ray crystallographic studies of 20G9 will allow thorough characterisation of the 20G9 active site and verify if the 20G9 Fv model generated by *AbM* is correct.

This project was badly affected by the problems encountered with expression of recombinant scFv antibody, although similar expression difficulties have been widely reported [Kipriyanov *et al*, 1997]. Not withstanding these problems, an attempt was made to understand the mechanism of this catalytic antibody based on a mutagenesis study guided by a 3-D model of 20G9. The project was successful insofar as it identified a residue that when eliminated, caused a total loss of activity.

APPENDIX

Synthesised oligonucleotides:

HindFor2: 5'-CTGATGTACCCAAGCTTATGAAATACCTATTGCCTAC -3'
Length: 37mer T_m: 54°C
Forward oligonucleotide which is used to introduce the HindIII restriction site. Can be used in combination with M-5 oligonucleotides to produce a forward overlap.

NotIback2: 5'-GAAGGAAAAAAGCGGCCGCTGCAGAGACAGTGACCAGAGT-3'
Length: 40mer T_m: 68°C
Backward oligonucleotide used to introduce the NotI restriction site. Can also be used in combination with M-3 oligonucleotides to produce a reverse overlap.

Link Seq. H: 5'-GAGGGTAAATCCTCAGG-3'
Length: 17mer T_m: 52 °C
Forward oligonucleotide that anneals to the linker sequence of 20G9 scFv allowing sequencing of the V_H region of 20G9.

Link Seq. L: 5'-TTTGGATTCGGAGCCAG-3'
Length: 17mer T_m: 52 °C
Reverse oligonucleotide that anneals to the linker sequence of 20G9 scFv allowing sequencing of the V_L region of 20G9.

LS1: 5'-CGGCTCGTATAATGTGTGGAAT-3'
Length: 22mer T_m: 68°C
Forward oligonucleotide that anneals to the vectors pKK233-2, pTrc99A upstream of the insert allowing PCR amplification and sequencing.

LS2: 5'-ACCGCTTCTGCGTTCTGATTT-3'
Length: 21mer T_m: 62°C
Reverse oligonucleotide that anneals to the vectors pKK233-2, pTrc99A downstream of the insert allowing PCR amplification and sequencing.

MY91A5: 5'-TATTCTAGAGCTTGCATACA-3'
Length: 20mer T_m: 54°C

MY91A3: 5'-TGTATGCAAGCTCTAGAATA-3'
Length: 20mer T_m: 54°C

MY94A5 5'-CGTGTACGGAGCTTCTAGATA-3'
Length: 21mer T_m: 62°C

MY94A3 5'-TATCTAGAAGCTCCGTACACG-3'
Length: 21mer T_m: 62°C

MY96A5 5'-TCCGAACGTGGCCGGATATTC-3'
Length: 21mer T_m: 66°C

MY96A3 5'-GAATATCCGGCCACGTTCGGA-3'
Length: 21mer T_m: 66°C

MR50A5 5'-AGGATGAATAGCGCCAATCCA-3'
Length: 21mer T_m: 52°C

MR50A3 5'-TGGATTGGCGCTATTCATCCT-3'
Length: 21mer T_m: 52°C

MR56A5 5'-TAACCTAGTAGCATTATCGGA-3'
Length: 21mer T_m: 48°C

MR56A3 5'-TCCGATAATGCTACTAGGTTA-3'
Length: 21mer T_m: 48°C

MR58A5 5'-CTGATTTAAAGCAGTTCTATT-3'
Length: 21mer T_m: 44°C

MR58A3 5'-CCGATAATAGAACTGCTTTAAAT-3'
Length: 23mer T_m: 48°C

pelB20G9His5 scFv nucleotide sequence

```
 M   K   Y   L   L   P   T   A   A   A   G   L   L   L
ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTG
           10          20          30          40

 L   A   A   Q   P   A   M   A   D   I   V   M   T   Q
CTA GCT GCC CAA CCA GCG ATG GCC GAT ATT GTG ATG ACT CAG
        50          60          70          80

 A   A   P   S   V   I   V   T   P   G   E   S   V   S
GCT GCA CCC TCT GTG ATT GTC ACT CCT GGA GAG TCA GTA TCC
     90          100         110         120

 I   S   C   R   S   S   K   S   L   L   H   S   Y   G
ATC TCC TGC AGG TCT AGT AAG AGT CTC CTG CAT AGT TAT GGC
  130         140         150         160

 N   T   Y   L   Y   W   F   L   Q   R   P   G   Q   S
AAC ACT TAC TTG TAT TGG TTC CTG CAG AGG CCA GGC CAG TCT
170         180         190         200         210

 P   Q   V   L   I   Y   R   M   S   N   L   A   S   G
CCT CAG GTC CTG ATA TAT CGG ATG TCC AAC CTT GCC TCA GGA
          220         230         240         250

 V   P   D   R   F   S   G   S   G   S   G   T   A   F
GTC CCA GAC AGG TTC AGT GGC AGT GGG TCA GGA ACT GCT TTC
       260         270         290         300

 T   L   R   I   S   R   V   E   A   D   D   V   G   V
ACA CTG AGA ATC AGT AGA GTG GAG GCT GAC GAT GTG GGT GTT
    310         320         330         340

 Y   Y   C   M   Q   Y   L   E   Y   P   Y   T   F   G
TAT TAC TGT ATG CAA TAT CTA GAA TAT CCG TAC ACG TTC GGA
  350         360         370         380

 G   G   T   K   L   E   L   K   E   G   K   S   S   G
GGG GGG ACC AAG CTG GAG CTG AAA GAG GGT AAA TCC TCA GGA
390         400         410         420         430

 S   G   S   E   S   K   V   D   E   V   Q   L   H   E
TCT GGC TCC GAA TCC AAA GTC GAC GAG GTG CAG CTG CAC GAG
          440         450         460         470

 P   G   A   A   L   V   R   P   H   A   S   V   M   L
CCT GGG GCT GCA CTG GTA AGG CCT GGA GCT TCA GTG ATG CTG
       480         490         500         510
```

```
 P   C   R   A   S   G   Y   S   F   T   S   F   W   M
CCC TGC AGG GCT TCT GGC TAC TCC TTC ACC AGC TTC TGG ATG
    520           530           540           550

 N   W   V   R   Q   R   P   G   Q   G   L   E   W   I
AAC TGG GTG AGG CAG AGG CCT GGA CAA GGC CTT GAG TGG ATT
  560           570           580           590

 G   R   I   H   P   S   D   N   R   T   R   L   N   Q
GGC AGG ATT CAT CCT TCC GAT AAT AGA ACT AGG TTA AAT CAG
600           610           620           630           640

 K   F   K   D   K   A   T   L   T   V   D   K   S   S
AAG TTC AAG GAC AAG GCC ACA TTG ACT GTA GAC AAA TCC TCC
          650           660           670           680

 S   T   A   Y   M   Q   L   I   S   P   T   S   E   D
AGC ACA GCC TAC ATG CAA CTC ATC AGT CCG ACC TCT GAG GAC
      690           700           710           720

 S   E   V   Y   Y   C   A   R   E   G   G   A   Y   W
TCT GAG GTC TAT TAC TGT GCA AGA GAG GGC GGG GCT TAC TGG
    730           740           750           760

 H   Q   H   T   L   V   T   V   S   A   H   H   H   H
GGC CAA GGG ACT CTG GTC ACT GTC TCT GCA CAC CAT CAC CAT
  770           780           790           800

 H   *   *
CAC TAA TGA
810
```

pUC119-20G9-His6*mycXbaI* scFv nucleotide sequence

```
 M   K   Y   L   L   P   T   A   A   A   G   L   L   L
ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTG
           10          20          30          40

 L   A   A   Q   P   A   M   A   D   I   V   M   T   Q
CTA GCT GCC CAA CCA GCG ATG GCC GAT ATT GTG ATG ACT CAG
        50          60          70          80

 A   A   P   S   V   I   V   T   P   G   E   S   V   S
GCT GCA CCC TCT GTG ATT GTC ACT CCT GGA GAG TCA GTA TCC
     90          100         110         120

 I   S   C   R   S   S   K   S   L   L   H   S   Y   G
ATC TCC TGC AGG TCT AGT AAG AGT CTC CTG CAT AGT TAT GGC
  130         140         150         160

 N   T   Y   L   Y   W   F   L   Q   R   P   G   Q   S
AAC ACT TAC TTG TAT TGG TTC CTG CAG AGG CCA GGC CAG TCT
170         180         190         200         210

 P   Q   V   L   I   Y   R   M   S   N   L   A   S   G
CCT CAG GTC CTG ATA TAT CGG ATG TCC AAC CTT GCC TCA GGA
          220         230         240         250

 V   P   D   R   F   S   G   S   G   S   G   T   A   F
GTC CCA GAC AGG TTC AGT GGC AGT GGG TCA GGA ACT GCT TTC
       260         270         290         300

 T   L   R   I   S   R   V   E   A   D   D   V   G   V
ACA CTG AGA ATC AGT AGA GTG GAG GCT GAC GAT GTG GGT GTT
    310         320         330         340

 Y   Y   C   M   Q   Y   L   E   Y   P   Y   T   F   G
TAT TAC TGT ATG CAA TAT CTA GAA TAT CCG TAC ACG TTC GGA
  350         360         370         380

 G   G   T   K   L   E   L   K   E   G   K   S   S   G
GGG GGG ACC AAG CTG GAG CTG AAA GAG GGT AAA TCC TCA GGA
390         400         410         420         430

 S   G   S   E   S   K   V   D   E   V   Q   L   H   E
TCT GGC TCC GAA TCC AAA GTC GAC GAG GTG CAG CTG CAC GAG
          440         450         460         470

 P   G   A   A   L   V   R   P   H   A   S   V   M   L
CCT GGG GCT GCA CTG GTA AGG CCT GGA GCT TCA GTG ATG CTG
       480         490         500         510
```

```
 P   C   R   A   S   G   Y   S   F   T   S   F   W   M
CCC TGC AGG GCT TCT GGC TAC TCC TTC ACC AGC TTC TGG ATG
    520           530           540           550

 N   W   V   R   Q   R   P   G   Q   G   L   E   W   I
AAC TGG GTG AGG CAG AGG CCT GGA CAA GGC CTT GAG TGG ATT
  560           570           580           590

 G   R   I   H   P   S   D   N   R   T   R   L   N   Q
GGC AGG ATT CAT CCT TCC GAT AAT AGA ACT AGG TTA AAT CAG
600           610           620           630           640

 K   F   K   D   K   A   T   L   T   V   D   K   S   S
AAG TTC AAG GAC AAG GCC ACA TTG ACT GTA GAC AAA TCC TCC
          650           660           670           680

 S   T   A   Y   M   Q   L   I   S   P   T   S   E   D
AGC ACA GCC TAC ATG CAA CTC ATC AGT CCG ACC TCT GAG GAC
      690           700           710           720

 S   E   V   Y   Y   C   A   R   E   G   G   A   Y   W
TCT GAG GTC TAT TAC TGT GCA AGA GAG GGC GGG GCT TAC TGG
    730           740           750           760

 H   Q   H   T   L   V   T   V   S   A
GGC CAA GGG ACT CTG GTC ACT GTC TCT GCA
  770           780           790
```

REFERENCES

Albery, W.J. & Knowles, J.R. (1977) Efficiency and evolution of enzymatic catalysis. *Angew. Chem. Int. Ed. Engl.*, **16**, 285-293.

Al-Lazikani, B., Lesk, A.M. & Chothia, C. (1997) Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology,* **273**, 4, 927-948.

Amann, E. and Brosius, J., (1985) Plasmid vectors for the regulated high level expression of eukaryotic genes in *Escherichia coli. Gene,* **40**, 183

Amann, E., Brosius, J. & Ptashne, M. (1983) The vectors bearing a hybrid *trp-lac* promotor useful for regulated expression of cloned genes in *Escherichia coli. Gene,* **25**, 167-178.

Amman, E., Ochs, B. & Abel, K.J. (1988) Tightly regulated *tac* promotor vectors for the expression of unfused and fused proteins in *Escherichia coli. Gene*, **69**, 301-315.

Baca, M., Scanlan, T.S. & Stephenson, R.C. & Wells, J.A. (1997) Phage display of a catalytic antibody to optimise affinity for TSA binding. *Proceedings of the National Academy of Sciences of the United States of America.,* **94**, 19, 10063-10068.

Barbas, C.F., Bain, J.D., Hoekstra, D.M. and Lerner, R.A. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proceedings of the National Academy of Sciences of the United States of America.*, **89**, 4457-4465.

Barre, S., Greenberg, A.S., Flajnik, M.F. & Chothia, C. (1994) Structural conservation of hypervariable regions in immunoglobulins evolution. *Nature Structural Biology,* **1**, 12, 915-920.

Bebbington, C.R. (1991) Expression of antibody genes in non-lymphoid mammalian cells. *METHODS: A Companion to Methods in Enzymology,* **2**, 2 136

Benkovic, S.J. (1996) The key is in the pocket. *Nature,* **383**, 23-24.

Benkovic, S.J., Adams, J.A., Borders, C.L., Janda, K.D. and Lerner, R.A. (1990) The enzymic nature of antibody catalysis: Development of multistep kinetic processing. *Science*, **250**, 1135-1139.

Benvenuto, E. and Tavladoraki, P. (1995) Immunotherapy of plant viral diseases. *Trends in Microbiology*, **3**, 272-274

Better, M., Chang, C.P., Robinson, R. R. and Horwitz, A. H. (1988) Escherichia coli secretion of an active chimeric antibody fragment. *Science*, **240**, 1041-1043

Bhat, T.N., Bentley, G.A., Boulot, G., Greene, M.I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F.P., Mariuzza, R.A. and Poljak, R.J. (1994) Bound water molecules and conformational stabilisation help mediate an antigen-antibody association. *Proceedings of the National Academy of Sciences of the United States of America.*, **91**, 1089-1093.

Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.W., Lee, T., Pope, S.H., Riordan, G.S., and Whitlow, M. (1988) Single-chain antigen-binding proteins., *Science,* **242**, 423-426.

Blackburn, G.F., Talley, D.B., Booth, P.M., Durfor., C.N., Martin, M.T., Napper, A.D. & Rees, A.R. (1990) Potentiometric Biosensor employing catalytic antibodies as the molecular recognition element. *Analytical Chemistry*, **62**, 2211-2216.

Blundell, T.L. and Sternberg, M.J.E. (1985) Computer-aided design in protein engineering. *Trends in Biotechnology*, **3**, 228-235

Bowdish, K., Tang, Y., Hicks, J.B. and Hilvert, D. (1991) Yeast expression of a catalytic antibodywith chorismate mutase activity. *Journal of Biological Chemistry*, **266,** 11901-11908

Braden, B. C., Souchon, J., Eisele, J.-L., Bentley, G.A., Bhat, T.N., Navaza, J., and Poljak, R.J. (1994) Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *Journal of Molecular Biology,* **243**, 767-781.

Browne, W., North, A., Phillips, D., Brew, K., Vanaman, T. and Hill, R. (1969) A possible three dimensional structure of bovine α-lactalbumin based on that of hen's egg lysozyme. *Journal of Molecular Biology*, **42**, 65-68.

Bruccoleri, R.E. and Karplus, M., (1987) Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers*, **26**, 137-168.

Buchner, J., Pastan, I. and Brinkmann, U. (1992) A method for increasing the yield of properly folded recombinant fusion proteins; single chain immunotoxins form renaturation of bacterial inclusion bodies. *Analytical Biochemistry,* **205**, 264.

Bullock, W.O., Fernandez, J.M. & Short, J.M. (1987) XL1-Blue: A high efficiency plasmid transforming RECA *Escherichia coli* strain with β-galactosidase selection. *BioTechniques,* **5**, 376-378

Calabi, F. & Neuberger, M.S. (eds) (1987) *Molecular Genetics of Immunoglobulin.* Elsevier Science Publishers, Amsterdam.

Chang, C.Y., Whitaker, P.B., Tabernero, L., Einspahr, H., Workman, L., Benjamin, D.C., and Sheriff, S. (1994) Crystallisation and preliminary x-ray analysis of an ant-staphylococcal nuclease-staphylococcal nuclease complex and of a second anti-staphylococcal antibody. *Journal of Molecular Biology*, **239**, 154-157.

Charbonnier, J. B., Carpenter, E., Gigant, B., Golinelli-Pimpaneau, B., Eshhar, Z., Green, B.S. & Knossow, M. (1995) Crystal structure of the complex of a catalytic antibody Fab fragment with a transition-state analog: structural similarities in esterase-like catalytic antibodies. *Proceedings of the National Academy of Sciences of the United States of America.*, ***92***, 11721-11725.

Chothia, C. and Lesk, A.M. (1987) Canonical structures for the hypervariable regions of immunoglobulins *Journal of Molecular Biology,* **196**, 901-917.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D.R., Tulip, W.R., Colman, P.M. & Poljak, R.J. (1989) Conformations of immunoglobulin hypervariable regions. *Nature,* **343**, 877-883.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Amit, A., Mariuzza, R., Phillips,S. and Poljak, R., (1986) The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science*, **233**, 755-758.

Collet, T.A., Roben, P., O'Kennedy, R., Barbas, C.F., Burton, D.R. and Lerner, R.A. (1992) A binary plasmid system for shuffling combinatorial antibody libraries. *Proceedings of the National Academy of Sciences of the United States of America.*, **89**, 10026-10034.

Colcher, D., Bird, R. and Roselli, M. (1990). In vivo tumour targeting of a recombinant single-chain antigen-binding protein. *Journal of the National Cancer Institute,* **82**, 1191-1197.

Colman, P.M., (1988) Structure of antibody antigen complexes - implications for immune recognition. *Advances in Immunology.*, **43**,99,132.

Cornish-Bowden, A. & Wharton, C.W. (1988) *Enzyme Kinetics*, IRL Press, Oxford, UK.

Covell, D.G., and Mallqvist, A., (1997) Analysis of protein-protein interactions and the effects of amino-acid mutations on their energetics. The importance of water molecules in the binding epitope. *Journal of Molecular Biology,* **269**, 2, 281-297.

Darsley, M.J. & Rees, A.R. (1985) Three distinct epitopes within the loop region of hen egg lysozyme defined with monoclonal antibodies. *EMBO Journal*, **4**, 2, 383-392.

Darsley, M.J. & Rees, A.R. (1985) Nucleotide sequences of five anti-lysozyme monoclonal antibodies. *EMBO Journal*, **4**, 2, 393-398.

Darveau, R.P., Somerville, J.E., Fell, H.P. (1992) Expression of antibody fragments in *Escherichia coli*. *Journal of Clinical Immunoassay*, **15**, 25-29.

Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T. (1988) *Proteins: Structure, Function and Genetics*, **4**, 31

Davies, D.R. and Padlan, E.A. (1990) Antibody-antigen complexes. *Annual Review of Biochemistry*, **59**, 439-473.

De la Paz, P., Sutton, B.J., Darsley, M.J. & Rees, A.R. (1986) Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO Journal*, **5**, 2, 415-425.

Deonarain, M.P., Rowlinson-Busza, G., George, A.J.T. and Epenetos, A.A. (1997) Re-designed anti-human placental alkaline phosphotase single-chain Fv: soluble expression, characterisation and in vivo tumour targeting. *Protein Engineering,* **10**, 1, 89-98.

Durfor, C.N., Bolin, R.J., Sugasawara, R.J. and Massey R.J. (1988) Antibody Catalysis in Reverse Micelles. *Journal of the American Chemical Society*, 1988. **110**. 8713-8714.

Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, M. (1985) Isolation of monoclonal antibodies specific for human *c-myc* proto-oncogene product. *Molecular and Cellular Biology,* **5**, 3610-3616.

Erlanger, B. (1980). The preparation of antigenic hapten-carrier conjugates: a survey. *Methods in Enzymology,* **70**, 85-103.

Fersht, A., (1985) The basic equations of enzyme kinetics. *Enzyme Structure and Mechanism*. (2nd edn) W.H.Freeman & Co., New York. pp 98-107.

Fersht, A.R. & Knill-Jones, J.W. (1981) DNA polymerase accuracty and spontaneous mutation rates: frequencies of purine-purine, purine-pyrimidine and pyrimidine-pyrimidine mismatches during DNA replication. *Proceedings of the National Academy of Sciences of the United States of America.*, **78**, 4251-4255.

Field, H., Yarranton, G.T. & Rees, A.R. (1989) Expression of mouse immunoglobulin light and heavy chain variable regions in *Escherichia coli* and reconstitution of antigen binding activity. *Protein Engineering*, **3**, 7, 641-647.

Foote, J. & Winter, G. (1992) Antibody framwork residues affecting the conformation of the hypervariable loops. *Journal of Molecular Biology*, **224**, 2, 487-499

Fraipont, C., Adam, M., Nguyen-Disteche, M., Keck, W., Van Beeumen, J., Ayala, J.A., Granier, B., Hara, H. & Ghuysen, J.M. (1994) Engineering and overexpression of periplasmic forms of the penicillin-binding protein 3 of *Escherichia coli*. *Biochemical Journal,* **298**, 189-195.

French, D.L., Laskov, R. & Scharff, M.D. (1989) The role of somatic hypermutation in the generation of antibody diversity. *Science*, **244**, 1152-1157.

Gallacher, G. (1993). Polyclonal catalytic antibodies. *Biochemical Society Transactions*, **21**, 1087-1090.

Gibson, T.J. (1984) Studies of the Epstein-Barr virus Genome. *PhD Thesis*. Cambridge University, UK.

Gigant, B., Charbonnier, J.B., Eshhar Z., Green, B.S. and Knossow, M. (1997) X-ray structures of a hydrolytic antibody and of complexes elucidate catalytic pathway from substrate binding and transition state stabilisation through water attack and product release. *Proceedings of the National Academy of Sciences of the United States of America.,* **94,** 15, 7857-7861

Gillam, S., Astell, C.R. & Smith, M. (1980) Site-specific mutagenesis using oligo deoxyribonucleotides: isolation of a phenotypically silent ϕx174 mutant, with a specific nucleotide deletion at very high efficiency. *Gene,* **12**, 129-37.

Gilliland, L.K., Norris, N.A., Marquardt, H., Tsu, T.T., Hayden, M.S., Neubauer, M.G., Yelton, D.E., Mittler, R.S. & Ledbetter, J.A. (1996) Rapid and reliable cloning of antibody variable regions and generation of recombinant single chain antibody fragments. *Tissue Antigens*, **47**, 1-20.

Glockshuber, R., Malia, M., Pfitzinger, I. and Pluckthun, A. (1990) A comparison of strategies to stabilise immunoglobulin Fv-fragments. *Biochemistry,* **29**, 1362.

Golinelli-Pimpaneau, B., Gigant, B., Bizebard, T., Navaza, J., Saludjian, P., Zemel, R., Tawfik, D.K., Eshhar, Z., Green, B.S. & Knossow, M. (1994) Crystal structure of a catalytic antibody F_{ab} with esterase-like activity. *Structure*, **2**. 175-183.

Gorman, S.D. & Clark, M.R. (1990) Humanisation of monoclonal antibodies for therapy. *Seminars in Immunology,* **2**, 6, 457-466.

Gouveneur, V.E., Houk, K.N., de Pascual-Teresa, B., Beno, B., Janda, K.D., Lerner, R.A. (1993) Control of the exo and endo pathways of the Diels-Alder reaction by antibody catalysis. *Science*, **260**, 337-339.

Griest, R.E., Jeffrey, P.D., Taylor, G.L. and Rees, A.R. (1992) Crystallization of the Fab fragments of anti-peptide monoclonal antibody and a complex with peptide. *Journal of Molecular Biology,* **223**.

Griffiths, A.D., Williams, S.C., Hartley, O., Tomlinson, I.M., Waterhouse, P., Crosby, W.L., Kontermann, R.E., Jones, P.T., Low, N.M., Allison, T.J., Prospero, T.D., Hoogenboom, A.N., Harrison, J.L., Zaccolo, M., Gherardi, E. and Winter, G. (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO Journal*, **13**, 3245.

Gussow, D. & Clackson, T. (1989) Direct clone characterisation from plaques and colonies by the polymerase chain-reaction. *Nucleic Acids Research*, **17**, 10, 4000-4000.

Gussow, D., Ward, E.S., Griffiths, A.D., Jones, P.T. & Winter, G. (1989) Generation of binding activities from Escherichia coli. by expression of immunoglobulin variable domains. *Cold Spring Harbour Symp. Quant. Biol.*, **1**, 265-273.

Hanahan, D. (1983) Studies on transformation of the *Escherichia coli* with plasmids. *Journal of Molecular Biology,* **166,** 557

Harris, L.J., Skaletsky, E. & Mcpherson, A. (1995) Crystallization of intact monoclonal antibodies. *Proteins*, **23**, 2, 285-289

Haseman, C.A. & Capra, J.D. (1991) Mutational analysis of arsonate binding by a CRIA+ antibody. V_H and V_L junctional diversity are essential for binding activity. *Journal of Biological Chemistry*, **266**, 12, 7626-7632.

Hatt, J., Callahan, M. and Green, A. (1992) *Strategies,* **5**, 2

Haynes, M.R., Heine, A. & Wilson, I.A. (1996) Catalytic antibody structures: an early assessment. *Israel Journal of Chemistry*, **36**, 133-142.

Haynes, M.R., Stura, E.A., Hilvert, D. & Wilson, I.A. (1994) Routes to Catalysis: Structure of a catalytic antibody and comparison with its natural counterpart. *Science*, **263**, 646-653.

Higuchi, R., Krummel, B. & Saikir, K. (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments. Study of protein and DNA interactions. *Nucleic Acids Research,* **16**, 7351-67.

Hilvert, D., Carpenter, S.H., Nared, K.D., Auditor, M.T.M. (1988) Catalysis of concerted reactions by antibodies: the Claisen rearrangement. *Proceedings of the National Academy of Sciences of the United States of America.*, **85**, 14, 4953-4955.

Hilvert, D., Hill, K.W., Nared, K.D. and Auditor, M.M. (1989) Antibody catalysis of a Diels-Alder reaction. *Journal of the American Chemical Society*, **111**, 9261-9262.

Hilvert, D., Nared, K.D., (1988) A stereospecific Claisen rearrangement catalysed by an antibody. *Journal of the American Chemical Society,* 1988, **110**, 5593-5594.

Ho, S.N., Junt, H.D., Horton, R.M., Pulolen, J.K. & Pease, L.R.(1989) Site directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene,* **77**, 51-59

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent. *Bio/Technology,* 1321-1325.

Houk, N. (1996) Predicting Antibody Catalyst Selectivity from Optimum Binding of Catalytic Groups to a Hapten. *Journal of the American Chemical Society,* **118**, 9204-9205

Huston, J.S., Levison, D., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M.N., Ridge, R.J., Bruccoleri, R.E., Haber, E., Crea, R. and Oppermann, H. (1988) *Proc. Natl. Acad.Sci.*, **85**, 5879-5883.

Ikemura, T. (1981b) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. A proposal for a synonymous codon choice that is optimal for the *Escherichia coli* translational system. *Journal of Molecular Biology*, **151**, 389-409.

Iliades, P., Kortt, A.A. & Hudson, P.J. (1997) Triabodies - single chain Fv fragments without a linker form trivalent trimers. *FEBS Letters,* **409**, 3, 437-441.

Ishizuka, H., Hanamura, A., Inada, T. & Aiba (1994) Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the CRP gene. *EMBO Journal*, **13**, 13, 3077-3082.

Iverson, B.L., Lerner, R.A. (1989) Sequence-specific peptide cleavage catalysed by an antibody. *Science,* **243**, 1184-1188.

Jacobsen, J.R., Schultz, P.G. (1994) Antibody catalysis of peptide bond formation. *Proc. Natl. Acad.Sci.*, **91**, 5888-5892.

Janda, K.D., Scholoeder, D., Benkovic, S.J., Lerner, R.A. (1988) Induction of an antibody that catalyses the hydrolysis of an amide bond. *Science*, **241,** 1188-1191.

Jencks, W.P. (1969) *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York. pp288.

Jencks, W.P. (1966) *Current Aspects of Biochemical Energetics*, ed. Kaplan, N.O. & Kennedy, E.P. Academic Press, New York, pp273-298.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S & Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, 5th edit., U.S. Dept. of Health and Human Services, Washington, DC.

Kiefhaber, T., Rudolf, R., Kohler, H.-H and Buchner, J. (1991) Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. *Bio/Technology,* **9**, 825.

King, D.J., Turner, A., Farnsworth, A.P.H, Adair, J.R., Owens, R.J., Pedley, R.B., Baldock, D., Proudfoot, K.A., Lawson, A.D.G., Beeley, N.R.A., Millar, K., Millican, T.A., Boyce, B.A., Antoniw, P., Begent, R.H.J., Shochat, D. & Yarranton, G.T. (1994) Improved tumour targeting with chemically cross-linked recombinant antibody fragments. *Cancer Research*, **54**, 66176-6185.

Kipriyanov, S.M., Dubel, S., Breitling, F., Kontermann, R.E. and Little, M. (1994) Recombinant single-chain Fv fragments carrying C-terminal cysteine residues: production of bivalent and biotinylated miniantibodies. *Molecular Immunology,* **31**, 1047.

Kipriyanov, S.M., Moldenhauer, G. and Little, M. (1997) High level production of soluble single chain antibodies in small-scale *Escherichia coli* cultures. *Journal of Immunological Methods,* **200**, 69-77

Kipriyanov, S.M., Moldenhauer, Martin, A.C.R, Kipriyanov, O.A. and Little, M. (1997) Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bacterial secretion but not the affinity. *Protein Engineering,* **10**, 445-453

Kleinfield, R., Hardy, R.R., Tarlinton, D., Dangl, J., Herzenberg, L.A., Weigert, M. (1986) Recombination between an expressed immunoglobulin heavy chain gene and a germline variable gene segment in a β-cell lymphoma. *Nature,* **322**, 843

Knappik, A. & Pluckthun, A. (1995) Engineered turns of a recombinant antibody improve its *in vivo* folding. *Protein Engineering*, **8**, 81-89.

Kocks, C. & Rajowsky, K. (1989) Stable expression and somatic hypermutation of antibody V regions in β-cell development pathways. *Annual Review of Immunology*, **7**, 537-539.

Kohler, G. & Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature, **256***, 495-497.

Kramer, B., Kramer, W.& Fritz, H.J. (1984a) Different base/base mismatches are corrected with different efficiences by the methyl-directed DNA mismatch repair system of *E.coli. Cell,* **38**, 879-88

Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography,* **24**, 946-950.

Kustervan-Someren, M., Flipphi, M., de Graaf, L., Vanden Broek, H., Kester, H., Hinnen, A. & Vissor, J. (1992) Characterisation of the *Aspergillus niger pel*B gene:

Structure and regulation of expression. *Molecular and General Genetics*, **234**, 1, 113-120.

Landry, D.W., Zhao, K., Yang, G.X., Glickman, M., Georgiadis, T.M. (1993) Antibody-catalyzed degradation of cocaine. *Science*, **259,** 1899-1901.

Lescar, J., Riottot, M. M., Souchon, H., Chitarra, V., Bentley, G.A., Navazza, J., Alzari, P.M., and Poljak, R. J. (1993) Crystallisation, preliminary X-ray diffraction study, and crystal packing of a complex between anti-hen lysozyme antibody F9.13.7 and guinea-fowl lysozyme. *Proteins structure, Function and Genetics,* **15**, 209-212.

Lesk, A.M. & Chothia, C. (1982) Evolution of proteins formed by beta-sheets .2. The core of the immunoglobulin domains. *Journal of Molecular Biology*, **160**, 2, 325-342.

Mamalaki, A., Trakas, N. & Tzartos, S.J. (1993) Bacterial expression of a scFv antibody fragment which efficiently protects the acetylcholine receptor against a genetic modulation caused by myasthenic antibodies. *European Journal of Immunology,* **23**, 1839-1845.

Marquart, M., Deisenhofer, J., Huber, R. (1980) Crystollographic refinement and atomic models of the intact immunoglobulin molecule KOL and its antigen-binding fragment at 3.0**Å** and 1.9**Å** resolution. *Journal of Molelcuar Biology*, **141,** 369-391.

Martin, A.C.R., Cheetham, J.C. and Rees, A.R., (1989) Modelling antibody hypervariable loops: A combined algorithm. *Proceedings of the National Academy of Sciences of the United States of America.,* **86**, 9268

McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. (1990) Phage antibodies: filimentous displaying antibody variable domains. *Nature,* **348**, 552-554.

Michelson, B. J. (1995) Transformation of *Escherichia coli* increases 260 fold upon inactivation of T4 DNA ligase. *Analytical Biochemistry*, **225**, 172.

Mian, I.S., Bradwell, A.R. and Olsen, A.J., (1991) Structure, function and properties of antibody-binding sites. *Journal of Molecular Biology*, **217**, 1, 133-151.

Miller, J.H. & Albertini, A.M. (1983) Effects of surrounding sequence on the suppression of nonsense codons. *Journal of Molecular Biology*, **164**, 1, 59-71.

Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A.M. (1998) Conformation of the third hypervariable region in the V_H domain of immunoglobulins. *Journal of Molecular Biology,* **275**, 2, 269-294.

Needleman, S. and Wunsch, C. (1970) A general method applicaple to the search for similarity in the amino acid sequence of two proteins. *Journal of Molecular Biology*, **48**, 443-453.

Newton, S.I., (1729), (1960) Principia (*Mathematical Principles of Philosphosphy and his System of the World*). University of California Press, Berkley California, fourth printing edition.

Nishima, T., Tsuji, A., Fukushima, D.D. (1974) Site of conjugation of BSA to corticosteroid hormones and specificity of antibodies. *Steroids*. **24**, 861-74.

Old, R.W. & Primrose, S.B. (1994) *Principles of Gene Manipulation: An introduction to genetic engineering.* Fifth edition. Blackwell Scientific Publications.

Orlandi, R., Gussow, D.H., Jones, P.T. & Winter, G. (1989) Cloning immunglobulin variable domains for expression by the polymerase chain-reaction. *Proceedings of the National Academy of Sciences of the United States of America.*, **86**, 10, 3833-3837.

Owen, M., Gandecha, A., Cockburn, B., & Whitelam, G. (1992) Synthesis of a functional anti-phytochrome single chain-Fv protein in transgenic tobacco. *Bio/Technology,* **10**, 790-797.

Padlan, E., Davies, D., Pecht, I., Givol, D. and Wright, C., (1976) Model building studies of antigen-binding sites: the hapten-binding site of MOPc-315. *Cold Spring Harbour Symposium Quantitative Biochemistry*, **41**,627-637

Padlan, E. (1990) On the nature of antibody combining sites: Unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins: Structure, Function and Genetics,* **7**,112-124.

Palmer, T. (1995) *Understanding Enzymes.* An introduction to bioenergetics, catalysis and kinetics. Fourth Edition, Prentice Hall/Ellis Horwood. pp99.

Pantoliano, M.W., Bird, R.E., Johnson., S., Asel, E.D., Dodd, S.W., Wood, J.F. and Hardman, K.D. (1991) Conformational stability, folding and ligand binding affinity of scFv Ig fragments expressed in *Escherichia coli. Biochemistry*, **30**, 10117-10125

Patten, P.A., Gray, N.S., Yang, P.L., Marks, C.B., Wedemayer, G.J., Boniface, J.J., Stevens, R.C. & Schultz, P.G. (1996) The Immunological Evolution of Catalysis. *Science,* **271**, 1086-1091.

Pauling, L. (1948) Chemical achievement and hope for the future. *American Science*, **36**, 51-58

Pauling, L. (1946) Molecular architecture and biological reactions. *Chem. Eng. News*, **24**, 1375-1377.

Pedersen, J., Searle, S., Henry, A. and Rees, A.R. (1992) Antibody Modelling: Beyond Homology. *Immunomethods,* **1**, 126

Persoon, M.A.A., Caothien, R.H., Burton, D.R. (1991) Generation of diverse high affinity human monoclonal antibodies by repetoire cloning. *Proceedings of the National Academy of Sciences of the United States of America.,* **88,** 2432-2436.

Picken, R., Mazaitis, A., Maois, W., Rey, M. and Heyneker, H. (1983) Nucleotide sequence of the gene for heat stable enterotoxin II of *Escherichia coli. Infection and Immunity*, 269-275.

Pluckthun, A. (1991) Antibody engineering: Advances from the use of *Escherichia coli* expression systems. *Bio/Technology,* **9**, 545

Pluckthun, A. (1994) Antibodies from *Escherichia coli*: the pharmacology of monoclonal antibodies. (Rosenberg, M. & Moore, G.P. eds). *Handbook of experimental pharmacology*, 113, Springer verlag, Berlin. 269-315.

Pollack, S.J., Jacobs, J.W., Schultz, P.G. (1986) Selective chemical catalysis by an antibody. *Science,* **234**, 1570-1573.

Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature,* **258**, 598.

Proba, K., Liming, G. and Pluckthun, A. (1995) Functional antibody single-chain fragments from the cytoplasm of *Escherichia coli*: influence of thioredoxin reductase (TrxB). *Gene,* **159** (203-207).

Rajan, S.S., Ely, K.R., Abola, E.E., Wood, M.K., Colman, P.M., Athay, R.J. and Edmundson, A.B. (1983) *Molecular Immunology,* **20**, 787-799.

Raso, V. & Stollar, B.D. (1975) The antibody-enzyme analogy. Characterisation of antibodies to phosphopyridoxyl-tyrosine derivatives. *Biochemistry*, **14**, 584-591.

Rees, A.R, Searle, S.J., Henry, A.H., Whitelegg, N. and Pedersen, J. (1996) *Protein structure prediction: a practical approach.* Antibody combining sites: structure and prediction. (Ed. Sternberg, MJE). pp 141-170.

Reiter, Y. & Pastan, I. (1998) Recombinant Fv immunotoxins and Fv fragments as novel agents for cancer therapy and diagnosis. *Trends in Biotechnology,* **16**, 513-520

Rini, J.M., Schulzegahmen, U. and Wilson, I.A., (1992) Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*, **255**, 5047, 959-965.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd Ed. Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklens, S and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America.* **74**, 5463.

Sarma, V.R., Siverton, E.W., Davies, D.R. and Terry, W.D. (1971) The three-dimensional structure at 6A resolution of a human gamma G1 immunoglobulin molecule. *Journal of Biological Chemistry,* **246**, 3753-37659.

Scharf, S.J., Horn, G.T. & Erlich, H.A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science,* **233**, 1076-8.

Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C. and Davies, D.R., (1987) Three-dimensional structure of an antibody-antigen complex. *Proceedings of the National Academy of Sciences of the United States of America.*, **84**,8075-8079.

Shuster, A., Gololobov, G.V., Kvashuk, O.A., Bogomolova, A.E., Smirnov, I.V., Gabibov, A.G. (1992) DNA hydrolysing autoantibodies. *Science*, **256**, 665-667.

Skerra, A. and Pluckthun, A. (1988) Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli. Science,* **240**, 1038-1041.

Skerra, A. and Pluckthun, A. (1991) Secretion and in vivo folding of the F_{ab} fragment of the antibody McPC603 in *Escherichia coli*: influence of disulphides and *cis*-prolines. *Protein Eng*ineering, **4**, 971-979.

Smith, R.G., Martin, M.T., Sanchez, R. and Kenten, J.H. (1993) *Methods in Molecular Biololgy: Antibody Engineering Protocols.* Cloning and Bacterial Expression of an Esterolytic sFv. **51**, 297-317.

Smith, R.G. (1994) Personal communication.

Somerville, J.E., Goshorn, S.C., Fell, H.P. and Darveau, R.P. (1994) Bacterial aspects associated with the expression of a single-chain antibody fragment in *Escherichia coli. Applied Microbiology and Biotechnology,* **42**, 595-603.

Spada, S., Krebber, C. and Pluckthun, A. (1997) Selectively infective phage. *Biological Chemistry*, **378**, 6, 445-456.

Stahl, M., Goldie, B., Bourne, S. and Thomas, N.R. (1995) *Journal of the American Chemical Society,* **117**, 5164-5165.

Stewart, J.D., Roberts, V.A., Thomas, N.R., Getzoff, E.D. & Benkovic, S.J. (1994) Site-directed mutagenesis of a catalytic antibody: An arginine and a histidine residue play key roles. *Biochemistry*, **33**, 1994-2003.

Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E.H., Johnson, M.J., Itakura, K., Wallace, R.B. (1981) *Developmental Biology Using Purified Genes*, 1st Ed. D.Brown. Acad. Press, New York.

Tagami, H., Inada, T., Kunimura, T. & Aiba, H. (1995) Glucose lowers CRP levels resulting in repression of the lac operon in cells lacking cAMP. *Molecular Microbiology*, **17**, 2, 251-258.

Takagi, H., Morinaga, Y., M., Ikemura, H. and Inouye, M. (1988) Mutant subtilisin E with enhanced protease activity obtained by site-directed mutagenesis. *Journal of Biological Chemistry*, **263**, 36, 19592-19596.

Tang, Y., Hicks, J.B., Hilvert, D. (1991) *In vivo* catalysis of a metabolically essential reaction by an antibody. *Proceedings of the National Academy of Sciences of the United States of America.*, **88**, 8784-8786.

Tavladoraki, P., Benvenuto, E., Trinca, S., Demartinis, D., Cattaneo, A. & Galeffi, P. (1993) Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack. *Nature,* **366**, 6454, 469-472.

Thomas, N.R. (1994) Hapten design for the generation of catalytic antibodies. *Applied Biochemistry and Biotechnology*, **47**, 2-3, 345-72

Thomas, N.R. (1996) Catalytic antibodies - reaching adolescence? *Current Developments in Bioorganic Chemistry,* **13**, 6, 479-512.

Tonegawa, S., (1983) Somatic generation of antibody diversity. *Nature,* **302,** 575-81

Tormo, J., Blaas, D., Parry, N.R., Rowlands, D., Stuart, D. & Fita, I. (1994) Crystal structure of a human rhinovirus neutralising antibody complexed with a peptide derived from viral capsid protein VP2. *EMBO Journal,* **13**, 2247-2256.

Tramontano, A., Janda, K.D., Lerner, R.A. (1986) Catalytic antibodies. *Science*, **234**, 1566-1570.

Trukhan, M.E., Gorovits, R.L., Lebedeva, M.I., Lapidus, A.L. & Mashko, S.V. (1988) Effectiveness of expression of chloramphenicol acetyltransferase gene controlled by foreign regulatory regions in Escherichia coli. III. Relative effectiveness of cloned promoters of *Escherichia coli* and coliphages. *Journal of Molecular Biology,* **22**, 4, 1033-1044.

Tulip, W.R., Varghese, J.W., Webster, R.G., Laver, W.G. & Colman, P.M. (1992) Crystal-structures of two mutant neuraminidase antibody complexes with amino-acid substitutions in the interface. *Journal of Molecular Biology,* **227**,1,149-159.

Ulrich, H.D., Patten, P.A., Yang, P.L., Romesberg, F.E. & Shultz, P.G. (1995) Expression studies of catalytic antibodies. *Proceedings of the National Academy of Sciences of the United States of America.*, **92,** 11907-11911.

Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods in Enzymology*, **153**, 3-11.

Wang, D., Liao, J., Mitra, D., Akolkar, P., Gruezo, F. and Kabat, E. (1991) The repertoire of antibodies to a single antigenic determinant. *Molecular Immunology*, **28**, 12, 1387-1397

Webster, D.M., Henry, A.H. & Rees, A.R. (1994) Antibody-antigen interactions. *Current Opinion of Structural Biology,* **4**, 857-867.

Wedemayer, G.J., Wang, L.H., Patten, P.A., Schultz, P.G. & Stevens, R.C. (1997) Crystal structures of the free and liganded form of an esterolytic catalytic antibody. *Journal of Molecular Biology*, **268**, 390-400.

Whitlow, M. and Filpula, D. (1991) Single-chain Fv proteins and their fusion proteins. *Methods: Companion to Methods in Enzymology,* **2**, 97-105.

Williamson, R.A., Persson, M.A.A. & Burton, D.R. (1991) Expression of a human monoclonal anti-(Rhesus-D) Fab fragment in *Escherichia coli* with the use of bacteriophage-lambda vectors. *Biochemical Journal*, **277,** Jul., 561-563.

Wilson, I.A. & Stanfield, R.L. (1994) Antibody-antigen interactions. *Current Opinions in Structural Biology*, **4**, 857

Winter, G., Griffiths, A.D., Hawkins, R.E. & Hoogenboom, H.R. (1994) Making antibodies by phage display technology. *Annual Review of Immunology,* **12**, 433-455.

Winter, G. & Milstein, C. (1991) Man-made antibodies. *Nature,* **349**, 293-299

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S. & Decruz, E. (1989) Quantitive-evaluation of *Escherichia coli.* host strains for tolerance to cytosine methylation in plasmid and phage recombination. *Nucleic Acids Research,* **17**, 3469-3478

Wu, X.C., Ng, S.C., Near, R.I. and Wong, S.L. (1993) Efficient production of a functional single-chain antidigoxin antibody via an engineered *bacillus subtilis* expression-secretion system. *Bio/Technology*, **11**, 71

Wysocki LJ, Gefter, ML Margolies, MN. (1990) Parallel evolution of antibody variable regions by somatic processes - consecutive shared somatic alterations in V_H genes expressed by independently generated hybridomas apparently acquired by point mutation and selection rather than by gene conversion. *Jouranl of Experimental Medicine,* **172** 315-323.

Yanisch-Perron, C., Viera, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains nucleotide sequences of the m13MP18 & pUC19 vectors. *Gene,* **33**, 103-119

Yip, T.T., Nakgawa, Y. and Porath, Y. (1989) Evaluation of the interaction of peptides with Cu(II), Ni(II) and Zn(II) by high performance immobilised metal ion affinity chromatography. *Analalytical Biochemistry,* **168**, 75.

Zdanov, A., Li, Y., Bundle, D.R., Deng, S-J., macKenzie, C.R., Narang, S.A., Young, N.M. and Cygler, M. (1994) *Proc. Natl. Acad, Sci.*, **91**, 6423-6427.

Zoller, M.J. & Smith, M. (1983) Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. *Methods in Enzymology*, **100**, 468-500.

Zhou, G.W., Guo, J., Huang, W., Fletterick, R.J. & Scanlan, T.S. (1994) Crystal structure of a catalytic antibody with a serine protease active site. *Science*, **265**, 1059-1064.